2016

Compromise, extremism, and guilt

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

COMPROMISE, EXTREMISM, AND GUILT

by

ALEXANDER T. POTERACK

B.A., New York University, 2010

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2016

Approved by

First Reader

Barton L. Lipman, PhD
Professor of Economics

Second Reader

Jawwad Noor
Associate Professor of Economics

Third Reader

Dilip Mookherjee
Professor of Economics

COMPROMISE, EXTREMISM, AND GUILT

ALEXANDER T. POTERACK

Boston University, Graduate School of Arts and Sciences, 2016

Major Professor: Barton L. Lipman, PhD
Professor of Economics

ABSTRACT

This dissertation is a study of non-standard economic behavior. The first chapter concerns two widely observed violations of Independence of Irrelevant Alternatives, the Compromise and Attraction effects. I construct a novel method of representing them by reducing the context of a menu to a frame, encompassing the worst option along each attribute in the menu, and observing a collection of preferences indexed by frames. The agent behaves as though a good's attractiveness along each attribute is judged relative to the frame with declining marginal utility. This allows me to give a novel interpretation of the compromise and attraction effects: they are consistent with indifference curves rotating clockwise as the frame moves down, and counter-clockwise as it goes left. It also allows me to give a representation theorem showing the behavioral axioms associated with a utility representation taking a good and the frame as arguments.

The second chapter applies the representation from Chapter One to electoral politics. It shows that incorporating these preferences generates equilibria where extremist candidates enter plurality elections in order to attractively frame their preferred moderate candidate, even if the extremists have probability zero of obtaining office themselves. While such candidates are frequently observed in elections, and there are papers generating equilibria with centrist sure losers (including Solow (2015)), this is

the first paper generating equilibria with these extremist candidates without unusual assumptions on election rules, or non single-peaked preferences. This paper creates a four candidate equilibrium with two extremist sure loser candidates, each on the fringes of opinion.

The third chapter concerns the effect of guilt on preferences in the circumstance of gift giving. A decision maker who experiences guilt may receive an increase in surplus from a gift card allowing guilt-free indulgence, potentially beyond even the surplus she'd receive from an equivalent cash gift. This paper isolates the behavior of guilt avoidance by exploiting a multi-period setting which incorporates a distinction between the decision maker's preferences over what she'd receive, and what she would choose. A representation inspired by Kopylov (2009) is adapted to this setting, providing a representation theorem for these preferences.

Contents

1 The Compromise and Attraction Effects Through Frame Preferences **1**

1.1 Introduction 1

1.2 Literature Review 7

1.3 The Model 10

1.3.1 Primitives 10

1.3.2 Frames, Revealed Preferences, and fWARP 10

1.3.3 Properties 11

1.3.4 Axioms 14

1.3.5 Representation Theorem 16

1.3.6 Alternate Frame Definitions 24

1.4 Conclusion 27

2 Extremist Politics and the Preference for Compromise **29**

2.1 Introduction 29

2.2 Related Literature 32

2.3 Model 34

2.4 Multidimensional Citizen-Candidate 36

2.5 Compromise Effects 42

2.6 Conclusion 49

3 Giving the Gift of Guilt Avoidance **52**

3.1 Introduction 52

3.2 Literature Review . . . 57

3.3 The Model . . . 58

3.3.1 Primitives . . . 58

3.3.2 Representation . . . 60

3.3.3 Axioms . . . 62

3.4 Conclusion . . . 65

A Appendices to The Compromise and Attraction Effects Through Frame Preferences 67

A.1 The n-attribute Case . . . 67

A.2 Non-monotonic Attribute Preferences . . . 71

B Appendices to Extremist Politics and the Preference for Compromise 73

B.1 Proofs of Theorems . . . 73

B.1.1 Multidimensional Citizen-Candidate without Compromise . . . 73

B.1.2 Equilibria with Compromise Effects . . . 76

C Appendices to Giving the Gift of Guilt Avoidance 104

C.1 Proof of Representation Theorem . . . 104

C.1.1 Preliminaries . . . 104

C.1.2 Axioms Imply Representation . . . 105

C.1.3 Representation Implies Axioms . . . 109

Bibliography 111

Curriculum Vitae 113

List of Tables

2.1 Election view of voter a 45
2.2 Election view of voter a' 46

List of Figures

1·1 Compromise and Attraction Effects 2

1·2 Frame movement creates both effects 4

1·3 Curve Rotation 5

1·4 Continuity Violation 8

1·5 Curve Rotation 12

1·6 Compromise/Attraction Monotonicity 15

1·7 Defining v_1 and v_2 17

1·8 Demonstrating Continuity of v_1 18

1·9 Continuity in Both Arugments 20

1·10 Pairwise Continuity 21

1·11 *Compromise/Attraction Monotonicity* and Indifference Curve Slope . 22

1·12 Necessity of *Frame Continuity* 24

1·13 Desire for Alternate Frame Definition 25

1·14 Alternate Frame Definition Not Producing Attraction Effect 27

2·1 Curve Rotation 43

2·2 Space Translation 44

2·3 Asymmetric Framing Effects 51

List of Abbreviations

$\mathbb{R}^2$	....	the Real plane
$\mathbb{R}^n$	....	the n-dimensional Real hyperplane

Chapter 1

The Compromise and Attraction Effects Through Frame Preferences

1.1 Introduction

Among the well-documented violations of the standard choice axioms are effects of menu and context not included in canonical theory. In the standard model of decision making, agents are aware of the entire universe of goods, with a clear personal ranking of said goods. Such an agent should, therefore, be immune to any effects related to the inclusion or exclusion of a good from a menu. However, contra this standard model, there is mounting evidence[1] of such effects. Decision makers appear to judge alternatives relative to what *is* in front of them, rather than what *could be*.

One circumstance in which such effects arise is when goods are comparable along several distinct attributes. For example, computers can be compared in terms of memory and processor speed; televisions can be compared in terms of size and picture quality; cars can be compared in terms of gas mileage and cargo space. It is easy to determine which computer has a faster processor, or which television is bigger, or which car gets better gas mileage. What is difficult, however, is determining how much of one attribute to trade off for improvement in another. This difficulty is the essence of the effects studied in this paper; experimentally, it is observed that decision makers are influenced in their decision of how to trade off between attributes by the

[1] The effects studied in this paper were first identified by Huber et al. (1982) and Simonson (1989).

presence of information the standard model considers extraneous.

The notable examples of this phenomenon considered herein are the compromise and attraction effects, which are best explained by example. Consider a decision maker planning to purchase a laptop, choosing between laptop $\mathbf{x}$, with a 2.5 GHz processor and 4 GBs of RAM; laptop $\mathbf{y}$, with a 2 GHz processor and 6 GBs of RAM; and laptop $\mathbf{z}$, with a 1.8 GHz processor and 8 GBs of RAM. The compromise effect is when agents choose $\mathbf{x}$ out of the menu $\{\mathbf{x}, \mathbf{y}\}$, and $\mathbf{y}$, but not $\mathbf{x}$, from the menu $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. That is, the presence of laptop $\mathbf{z}$ switches their choice from $\mathbf{x}$ to $\mathbf{y}$, in violation of WARP, because the presence of $\mathbf{z}$ as a "more extreme" option makes $\mathbf{y}$ appear to be a desirable "compromise."

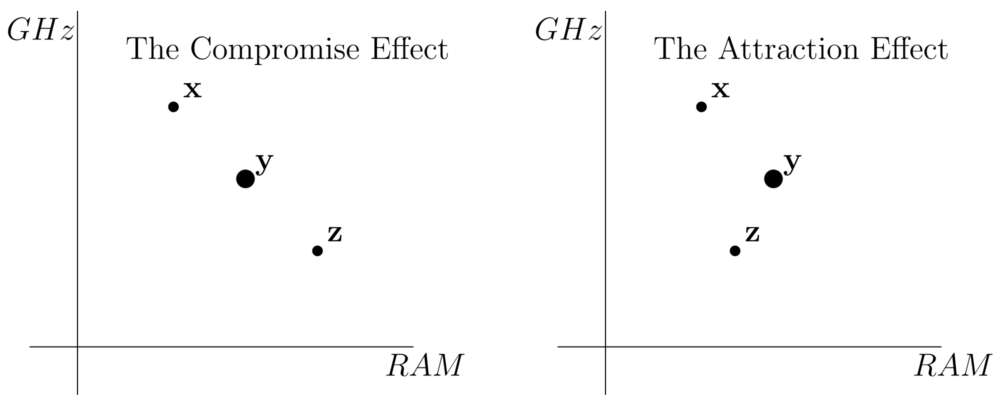

Figure 1·1: Graphical illustration of the effects

The "attraction effect" is similar. Suppose instead that laptop $\mathbf{z}$ represents a machine with a 1.8 GHz processor, but only 5 GBs of RAM. Some agents switch from choosing $\mathbf{x}$ out of $\{\mathbf{x}, \mathbf{y}\}$ to choosing $\mathbf{y}$ out of $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. $\mathbf{z}$ is no longer "extreme" and making $\mathbf{y}$ a "compromise;" it is, however, clearly dominated on both attributes by $\mathbf{y}$ and only $\mathbf{y}$, thereby making $\mathbf{y}$ seem more "attractive" as an easy choice. Notice that in Figure 1·1, the only difference between the effects is the horizontal position of $\mathbf{z}$.

In this example, decision makers must compare the marginal improvements of moving from a 2 GHz to a 2.5 GHz processor, or from 4 GBs of RAM to 6. The effects suggest that they draw inferences about how to judge these marginal improvements from the context created by the rest of the menu. One possible explanation for these phenomena is that this context is established by the worst good along each attribute.

Suppose the slowest processor speed establishes a "baseline" which all other processors are compared to, and the marginal utility of processing speed decreases moving further from the baseline. This is an idea introduced by Tversky and Kahneman (1991) in the context of loss aversion, which they call "diminishing sensitivity." In this case, the gain in moving from a 2 GHz processor to a 2.5 GHz processor would seem larger when the slowest processor is 2 GHz than it would in the presence of a 1.8 GHz processor. Intriguingly, both the compromise and attraction effects are consistent with this conceptualization.

This is the framework I will use. $\mathbb{R}^n$ is a set of goods[2]; the n dimensions represent n separate rankings over these goods. Each of these rankings is derived from an attribute along which goods are easily compared, such as memory or gas mileage. Decision makers make a choice from a subset of goods, which I will refer to as a "menu." There is a context created by this menu; following the example of Rubinstein and Salant (2008), I will call this context the "frame,"[3] and consider choice behavior which is standard for any fixed frame, but may demonstrate unusual effects when the frame is changed. More formally, there is a function mapping menus to frames, and when comparing choices from menus with the same frame, the decision makers' choices satisfy WARP. When comparing menus with different frames, WARP may be violated

The "baseline" interpretation illustrated in the laptop example naturally leads to

[2] In the body of the paper, I consider the preferences over $\mathbb{R}^2$; I expand the model to consider $\mathbb{R}^n$ in Appendix A.1.

[3] This could also be called a "reference;" it fits naturally with the literature on reference dependence.

a definition of the frame as the worst value for each attribute among goods in the menu (i.e., in the laptop example, it would be (1.8 GHz, 4 GB RAM) once laptop **z** is included). One attractive feature of this notion of frames is that the compromise and attraction effects both arise from the same source. Consider the following example:

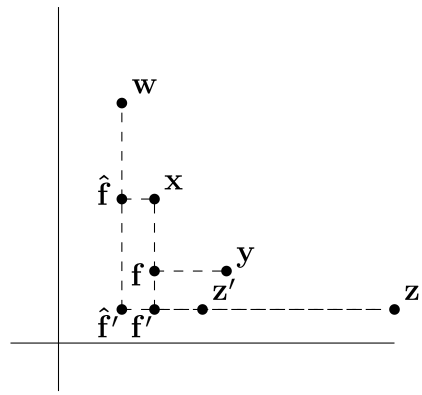

Figure 1·2: The compromise and attraction effects both create the same frame movement

Consider a choice correspondence $C(\cdot)$ such that $C(\{\mathbf{x}, \mathbf{y}\}) = \{\mathbf{x}, \mathbf{y}\}$. The compromise effect implies $C(\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = \{\mathbf{y}\}$. The attraction effect implies $C(\{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) = \{\mathbf{y}\}$. In both cases, the frame is lowered from $\mathbf{f}$ to $\mathbf{f}'$, and this change in frame is what changes the choice. The immediate observation is that lowering the frame makes $\mathbf{y}$ more appealing relative to $\mathbf{x}$.

However, it would be a mistake to conclude lowering the frame makes $\mathbf{x}$ less appealing relative to *all* other goods. When adding $\mathbf{z}$ to the menu $\{\mathbf{w}, \mathbf{x}\}$, it is still the case that the frame is lowered; however, now this causes $\mathbf{x}$ to be chosen over $\mathbf{w}$. In other words, compromise and attraction effects are consistent with a lowering of the frame making $\mathbf{x}$ *less* appealing relative to goods to its *right*, and *more* appealing relative to goods to its *left*.

This relationship is clearer when translated into terms of preferences. The assumption that WARP holds when the frame is held constant implies the existence of a collection of complete and transitive preferences indexed by frames. Denote this collection $\{\succsim^{\mathbf{f}}\}_{\mathbf{f}\in\mathbb{R}^2}$, where $\succsim^{\mathbf{f}}$ is the preference revealed by choices from menus with the frame $\mathbf{f}$. In the language of preferences, to say lowering the frame makes $\mathbf{x}$ less appealing relative to goods to its right and more appealing relative to goods to its left is to say it *rotates* the indifference curve associated with a given frame's revealed preference *clockwise* as the given frame is lowered. Similar analysis shows that moving the frame *left* rotates the curve *counterclockwise*[4].

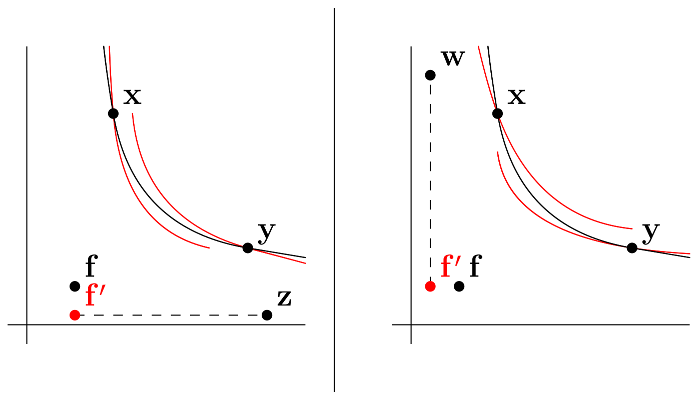

Figure 1·3: Indifference Curve Rotation in Response to Frame Changes

Thus, this choice of frame definition and description of indifference curve rotation combine to form a succinct and intuitive description of the compromise and attraction effects. Furthermore, there is a straightforward mathematical interpretation. The collection of preferences $\{\succsim^{\mathbf{f}}\}_{\mathbf{f}\in\mathbb{R}^2}$ can be represented by a function of the form $U(\mathbf{x},\mathbf{f})$. The rotation of indifference curves is a change in their slope. The slope of an indifference curve (holding the frame constant) is[5] $\frac{U_1}{U_2}$, and the desired rotation

[4]Consider repeating the analysis from the previous two paragraphs by adding $\mathbf{w}$ to the menu $\{\mathbf{x},\mathbf{y}\}$.

[5]Where $U_i \equiv \frac{\partial U}{\partial i}$, i.e., the partial derivative of U with respect to the ith argument.

with regard to the frame is equivalent to:

$$\frac{\partial}{\partial 3}\frac{U_1}{U_2} > 0$$
$$\frac{\partial}{\partial 4}\frac{U_1}{U_2} < 0$$

It will be shown that this is equivalent to $\frac{U_{13}}{U_{23}} > \frac{U_1}{U_2} > \frac{U_{14}}{U_{24}}$, a property I will call *Compromise/Attraction Rotation.* I will show in the body of the paper this is closely related to the notion of "diminishing sensitivity."

The compromise and attraction effects only arise in choices from menus of three or more goods. As such, I will assume that the preference revealed by choices from menus containing only two goods, hereafter referred to as the "pairwise preference," is complete on $\mathbb{R}^2$, transitive, and continuous. In addition to limiting the departures from the standard model to the behavior of interest, this provides structure for the frame preferences revealed by larger menus. An inherent limitation of the frame preference notion is that each $\succsim^{\mathbf{f}}$ is complete only over a subset of $\mathbb{R}^2$; certain goods cannot be in a set with frame $\mathbf{f}$. It is difficult to impose regularity on how preferences change with respect to the frame when there are certain goods over which preferences are only defined for one frame and not the other. By generating a preference over all of $\mathbb{R}^2$ from the frame preferences, this assumption allows for desired regularity, namely monotonicity and continuity in the frame.

After a brief literature review in Section 1.2, the paper proceeds in Section 1.3.1 by taking the choice function and space of goods as a primitive, then deriving the frame preferences in Section 1.3.2. From there, Section 1.3.3 enumerates a list of properties for the utility function representation of the collection of frame preferences, followed by the equivalent behavioral axioms in Section 1.3.4. A representation theorem is proven in Section 1.3.5, followed by a discussion of the frame definition in Section 1.3.6.

1.2 Literature Review

The attraction effect was first demonstrated experimentally by Huber et al. (1982), and the compromise effect was demonstrated first by Simonson (1989), whose paper also provided support for the attraction effect. These papers are strictly concerned with observing the effects; neither of them construct a representation incorporating these effects. An early model is Simonson and Tversky (1993), which (unlike mine) depends on context created by all elements in a menu, not just the worst along each attribute. Kivetz et al. (2004) analyze the Simonson and Tversky model, and two others, in terms of which best fit available experimental data. These papers propose several representations which demonstrate the effects; however, there is no axiomatization for any of these models, so their full behavioral implications remain unknown.

Ravid (2015) introduces a random choice procedure which allows for both effects. He shows this procedure can approximate a deterministic choice model if the deterministic model is equivalent to Simonson and Tversky (1993); it too depends on context created by all elements in a menu.

Ok et al. (2015) develop an axiomatized model which incorporates the attraction effect. Unlike my model, they endogenize the choice of reference (or frame). By construction, their model *cannot* represent the compromise effect. They have an axiom called “Reference Coherence” which essentially constructs a world where a good which “helps” another good in one context can never “harm” it in another context. However, the compromise effect helps a given good relative to some goods, but harms it relative to others. When $\mathbf{z}$ is down and to the right of $\mathbf{x}$, goods to $\mathbf{x}$’s right, between it and $\mathbf{z}$, appear to be compromises, and $\mathbf{z}$ can make them preferred over $\mathbf{x}$; i.e., $\mathbf{z}$ harms $\mathbf{x}$ relative to those goods. But $\mathbf{x}$ is between $\mathbf{z}$ and goods to the left of $\mathbf{x}$, so compared to those goods, $\mathbf{x}$ is the compromise, and $\mathbf{z}$ can make $\mathbf{x}$

preferred over goods to its left; i.e., $\mathbf{z}$ helps $\mathbf{x}$ relative to those goods.

Barbos (2010) also gives a representation theorem admitting the attraction effect but not the compromise effect. In that paper, goods are divided into exogenous categories, and the attraction effect privileges goods which have an inferior good within their category. Again, there isn't a clear way to fit the compromise effect into this framework.

The first paper that axiomatizes a choice representation which incorporates both the attraction and compromise effects is de Clippel and Eliaz (2012). They construct a representation based around a multiple selves bargaining game, where the ranking of a good along each attribute represents its attractiveness to a given self. In this representation, if the menu contains a good which neither self views as the worst option, it will be selected. This violates continuity, as illustrated in the example pictured:

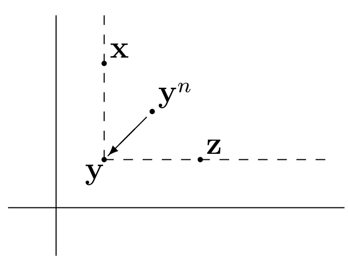

Figure 1·4: Violation of continuity in de Clippel Eliaz

If there is a sequence of goods converging to good $\mathbf{y}$ as pictured, every good in the sequence will be the sole choice out of a menu with $\mathbf{x}$ and $\mathbf{z}$, but $\mathbf{y}$ will not be chosen out of the menu $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. Furthermore, they require indifference between any $\mathbf{x}$ and $\mathbf{z}$ as depicted (neither dominant on both attributes), so there are no circumstances where adding an irrelevant alternative may switch a choice entirely from exclusively choosing one good to exclusively choosing a different good. My model has continuity, and allows for such switches.

The paper most closely related to the present work is Tserenjigmid (2015)[6]. While independently conceived and executed, both papers consider representations of choice behavior in which diminishing sensitivity generates both the compromise and attraction effects. The representations differ in that he characterizes the more specific functional form $g(u(x_1) - u(f_1)) + g(w(x_2) - w(f_2))$, with the frame of reference defined as it is in this paper, whereas I characterize a more general class of representations. Also, I show how to expand the set of goods considered from $\mathbb{R}^2$ to $\mathbb{R}^n$, an extension Tserenjigmid also notes is feasible for his representation.

The additively separable form of the representation is useful and tractable; it is employed by Poterack and Solow (2015) to study electoral politics. Tserenjigmid uses two axioms for this structure: the Thomsen condition and a translation invariance condition. In addition to generating the convenient functional form, these axioms also rule out some preference profiles that are potentially of interest. The Thomsen condition applies to all preferences revealed by choices, even those revealed by choices from two-good menus which cannot display either compromise or attraction effects. The translation invariance condition places restrictions on the magnitude of the effects as the menu is translated in good space. This poses difficulty for applications which wish to consider menu effects whose magnitudes vary in the good space. For example, in the earlier given description of a decision maker purchasing a laptop, the compromise and attraction effects seem very plausible. However, if this decision maker were instead considering various options for a network of servers costing hundreds of thousands of dollars, the decision would presumably be much more carefully considered and thus less influenced by these effects. Because the additively separable form allows for no interaction between the components of the frame, it is not well suited for modeling both kinds of decisions.

[6]The first draft of this paper was made available in 2014.

The general concept of choice influenced by a frame is related to a class of models explored by Rubinstein and Salant (2008). Rubinstein and Salant consider choice functions which take both the menu *and* a frame as arguments; despite my emphasis on frames, the choice correspondence in this paper only takes the menu as an argument, because the frame is a *function* of the menu.

1.3 The Model

1.3.1 Primitives

$\mathbb{R}^2$ represents a set of goods[7]. The two components of an element of $\mathbb{R}^2$ represent two "attributes" which are easily compared (e.g., RAM and processor speed).

The decision maker has a choice correspondence $C : \mathcal{S} \to \mathcal{S}$, where $\mathcal{S}$ is the collection of all proper compact[8] subsets of $\mathbb{R}^2$. For simplicity, assume that a good with a larger number in one component is preferred along the represented attribute. Mathematically, if $x_{-i} = y_{-i}$, then $\mathbf{x} \in C(\{\mathbf{x}, \mathbf{y}\})$ if and only if $x_i \geq y_i$. Define the attribute preferences $\gtrsim_1$ and $\gtrsim_2$ by $\mathbf{x} \succ_i \mathbf{y}$ if and only if $x_i \geq y_i$. These preferences are complete and transitive[9].

1.3.2 Frames, Revealed Preferences, and fWARP

The choice correspondence $C(\cdot)$ can be used to construct a collection of revealed frame preferences. First, the concept of a "frame" on a menu must be defined.

Define *"frame of S"* ($\mathbf{f}(S)$) by[10]

$$\mathbf{f}(S) \equiv \left(\min_{\mathbf{x} \in S} x_1, \min_{\mathbf{x} \in S} x_2 \right) \tag{1.3.1}$$

[7] The model can be extended to $\mathbb{R}^n$; details are found in appendix A.1.

[8] Compactness is used solely for the sake of a well-defined minimum.

[9] Monotonicity in the components can be relaxed; see appendix A.2.

[10] This definition of frame is very specific, but it will be shown in Section 1.3.6 that it can be relaxed.

Define a revealed frame preference, $\succsim^{\mathbf{f}}$ by

$$\mathbf{x} \succsim^{\mathbf{f}} \mathbf{y} \Leftrightarrow \exists\ S \in \mathcal{S} \text{ s.t. } \mathbf{x}, \mathbf{y} \in S, \mathbf{f}(S) = \mathbf{f}, \text{ and } \mathbf{x} \in C(S) \tag{1.3.2}$$

Note that by the definition of $\mathbf{f}(S)$, $\succsim^{\mathbf{f}}$ can only be defined for $\mathbf{x}, \mathbf{y} \in A^{\mathbf{f}}$, where

$$A^{\mathbf{f}} \equiv \left\{\mathbf{x} \in \mathbb{R}^2 | x_1 \geq f_1, x_2 \geq f_2\right\} \text{ (where } \mathbf{f} = (f_1, f_2)\text{)}.$$

There exists an $\succsim^{\mathbf{f}}$ for each $\mathbf{f} \in \mathbb{R}^2$. Collectively, they are $\{\succsim^{\mathbf{f}}\}_{\mathbf{f} \in \mathbb{R}^2}$.

Obviously, to study the compromise and attraction effects, WARP must be weakened. However, some regularity is still desired. It is given by the following axiom:

Axiom 1 (The Weak Axiom of Revealed Frame Preference (fWARP))**.** If for some $S \in \mathcal{S}$ with $\mathbf{x}, \mathbf{y} \in S$, $\mathbf{x} \in C(S)$, then for any $S' \in \mathcal{S}$ with $\mathbf{x}, \mathbf{y} \in S'$ and $\mathbf{f}(S) = \mathbf{f}(S')$, $\mathbf{y} \in C(S')$ implies $\mathbf{x} \in C(S')$.

This axiom is identical to standard WARP, save for the inclusion of the additional condition "$\mathbf{f}(S) = \mathbf{f}(S')$." This reflects the underlying idea the model: within a frame, preferences behave as expected. When the frame changes, the preferences do as well.

fWARP implies that $\{\succsim^{\mathbf{f}}\}_{\mathbf{f} \in \mathbb{R}^2}$ represents $C(\cdot)$ in the sense that

$$C(S) = \left\{\mathbf{x} \in S | \mathbf{x} \succsim^{\mathbf{f}(S)} \mathbf{y}\ \forall\ \mathbf{y} \in S\right\} \tag{1.3.3}$$

In light of this, I work with the frame preferences henceforth.

1.3.3 Properties

We will begin by considering the properties desirable in such a function. Firstly, properties to ensure tractability of the utility function are desired. Specifically, it should be monotone in the first two arguments, and continuous in all four arguments. Given that goods are assumed to be increasing in desirability in their components, the function must be strictly increasing in the first two arguments. No monotonicity restriction is placed on the third and fourth arguments because how the utility varies with $\mathbf{f}$, holding $\mathbf{x}$ constant, has no behavioral implications and is therefore irrelevant to the paper.; the relevant issue is how the *relative* values of the utilities for various goods change. The behavior of the representation with regard to the frame is entirely a statement about the behavior of the *collection* of preferences: the change when moving from $\succsim^{\mathbf{f}}$ to $\succsim^{\mathbf{f}'}$.

Property 1 (*Regularity*)**.** U is strictly increasing in the first two arguments and continuous in all four arguments.

Next is the property which gives the compromise and attraction effects, via the indifference curve rotation described in the introduction.

Property 2 (*Compromise/Attraction Rotation*)**.** [11] $\frac{U_{13}}{U_{23}} > \frac{U_1}{U_2} > \frac{U_{14}}{U_{24}}$

This is the property which most captures the nature of the effects. It is equivalent to the indifference curves rotating clockwise when the frame is lowered, and counter-clockwise as it is moved left. Consider the following example: We would expect, by

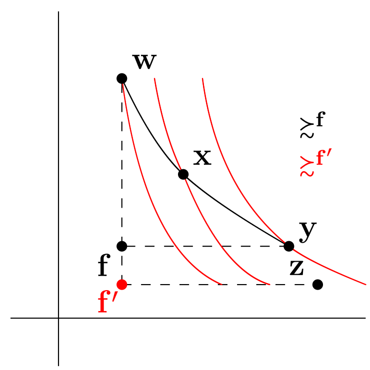

Figure 1·5: Indifference Curve Rotation in Response to Frame Changes

the compromise effect, the addition of $\mathbf{z}$ to the menu $\{\mathbf{w}, \mathbf{x}, \mathbf{y}\}$ to make $\mathbf{y}$ preferred

[11]The construction of this property suggests difficulty when derivatives are equal to zero or non-existent; this property can be expressed more generally to encompass these cases. Details can be found in the proof of the representation theorem.

to $\mathbf{x}$, but it would also make $\mathbf{x}$ preferred to $\mathbf{w}$. This corresponds to the clockwise rotation of indifference curves, in response to the frame being lowered from $\mathbf{f}$ to $\mathbf{f}'$. Similarly, the addition of $\mathbf{w}$ to the menu $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ would make $\mathbf{x}$ preferred to $\mathbf{y}$ and $\mathbf{y}$ preferred to $\mathbf{z}$. That corresponds to the counterclockwise rotation of indifference curves as the frame is moved left. Again, by the definition of frame used, this will also capture the attraction effect.

This is closely related to the property of "diminishing sensitivity," described by Tversky and Kahneman (1991). This property stipulates that the marginal value of a gain diminishes as it is further from the reference (or in this case, the frame). When the frame is lowered, the goods being compared are further from it. This means that a given advantage in the vertical attribute now has a lower value, and so a greater advantage is required to make up for a given disadvantage in the horizontal attribute—hence, the indifference curve must be steeper, which is achieved by its clockwise rotation. The same applies along the horizontal attribute when the frame is shifted left.

Finally, there are two more technical properties used for mathematical convenience. One guarantees transitivity among comparisons of pairs of elements; the other ensures that indifference curves cannot asymptote.

Property 3 (*Pairwise Transitivity*)**.** $U(x_1, x_2, x_1, y_2) = U(y_1, y_2, x_1, y_2)$
& $U(y_1, y_2, y_1, z_2) = U(z_1, z_2, y_1, z_2)$
$\Rightarrow U(x_1, x_2, x_1, z_2) = U(z_1, z_2, x_1, z_2)$[12]

Property 4 (*Non-asymptotic Indifference Curves*)**.** Given a frame $\mathbf{f}$, for each $\mathbf{x} \in A^{\mathbf{f}}$ there exists a $\mathbf{y} \in A^{\mathbf{f}}$ such that

$$U(x_1, x_2, f_1, f_2) = U(f_1, y_2, f_1, f_2) = U(y_1, f_2, f_1, f_2)$$

[12]This property exists to make Axiom 6, *Pairwise Weak Order*, necessary. If removed, *Pairwise Weak Order* is no longer necessary, but the rest of the axioms remain necessary, and *Pairwise Weak Order* remains sufficient.

1.3.4 Axioms

The crux of the idea is that preferences behave as usual when holding the frame fixed. As such, the first axiom says exactly that.

Axiom 1 (*Continuous Weak Order*). $\succsim^{\mathbf{f}}$ is complete on $A^{\mathbf{f}}$, transitive, and continuous, $\forall\ \mathbf{f} \in \mathbb{R}^2$.

There are two more axioms which apply within a fixed frame, and do not consider moving the frame.

Axiom 2 (*Simplicity*). $\mathbf{x} \succsim_1 \mathbf{y}$ and $\mathbf{x} \succsim_2 \mathbf{y} \Rightarrow \mathbf{x} \succsim^{\mathbf{f}} \mathbf{y}$ for all $\mathbf{f}$ such that $\mathbf{x}, \mathbf{y} \in A^{\mathbf{f}}$.

This is meant to capture the idea that choice is only difficult when the two attribute preferences disagree. If $\mathbf{x}$ is preferred to $\mathbf{y}$ on both attributes, it is always preferred to $\mathbf{y}$, regardless of context. It is equivalent to U being strictly increasing in the first two arguments.

Axiom 3 (*Substitutability*). Given $\mathbf{y}, \mathbf{f}$ such that $\mathbf{y} \in A^{\mathbf{f}}$, $\exists\ \mathbf{x}$ s.t. $x_1 = f_1$ and $\mathbf{y} \sim^{\mathbf{f}} \mathbf{x}$, and $\mathbf{z}$ s.t. $z_2 = f_2$ and $\mathbf{z} \sim^{\mathbf{f}} \mathbf{y}$.

This merely prevents asymptotic indifference curves; it is included for technical reasons, though it also captures the intuition that it's always possible to trade off between attributes. It is equivalent to Property 4.

The next axiom is the first to address how preferences change when the frame is moved. It is equivalent to Property 2; as such, it is my formal statement of the compromise and attraction effects:

Axiom 4 (*Compromise/Attraction Monotonicity*). Given $\mathbf{x} \sim^{\mathbf{f}} \mathbf{y}$ and $\mathbf{x} \succ_i \mathbf{y}$, then

1. $\mathbf{y} \succ^{(f_{-i}, f'_i)} \mathbf{x}\ \forall\ f'_i < f_i$,
2. $\mathbf{x} \succ^{(f_{-i}, f''_i)} \mathbf{y}\ \forall\ f''_i \in (f_i, y_i]$,
3. $\mathbf{x} \succ^{(f'_{-i}, f_i)} \mathbf{y}\ \forall\ f'_{-i} < f_{-i}$, and

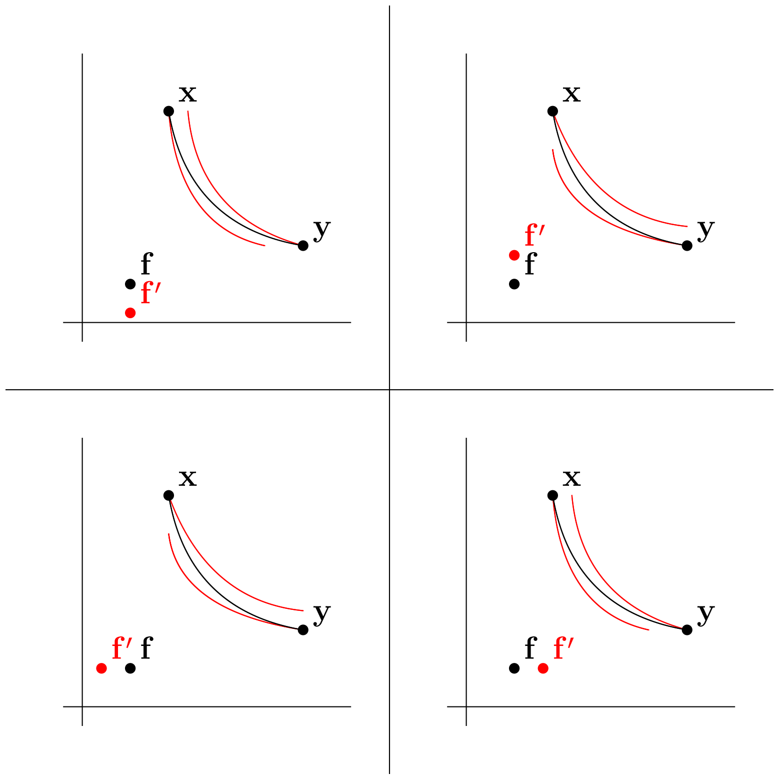

Figure 1·6: Compromise/Attraction Monotonicity, illustrated

4. $\mathbf{y} \succ^{(f''_{-i}, f_i)} \mathbf{x} \ \forall f''_{-i} \in (f_{-i}, x_{-i}]$.

A good which only lowers the frame cannot make $\mathbf{x}$ appear as a compromise relative to $\mathbf{y}$, nor be in position for the attraction effect to hold. Therefore, lowering the frame advantages $\mathbf{y}$. Similarly, a good which only moves the frame left cannot make $\mathbf{y}$ appear as a compromise relative to $\mathbf{x}$, nor be in position for the attraction effect to hold. Therefore, moving the frame left advantages $\mathbf{x}$.

The next axiom is a continuity axiom related to the behavior of the preferences as the frame changes. The same intuition which makes continuous preferences appealing also makes a continuity in how the preferences change with respect to the frame appealing. A small change in the frame should not make a sudden jump in the preferences. This intuition is captured by the following axiom:

Axiom 5 (*Frame Continuity*)**.** Given $\mathbf{f}$, $\mathbf{x}, \mathbf{y}$ such that $\mathbf{x} \succ^{\mathbf{f}} \mathbf{y}$, and f'_i such that

$\mathbf{y} \succ^{(f_{-i}, f_i')} \mathbf{x}$, then $\exists\ f_i''$ such that $\mathbf{x} \sim^{(f_{-i}, f_i'')} \mathbf{y}$.

As continuous preferences allow for continuous utility functions, *Frame Continuity* is a necessary condition for the representation to be continuous in $\mathbf{f}$. However, it is not a sufficient condition to guarantee continuity in $\mathbf{f}$, because each frame preference is incomplete on $\mathbb{R}^2$, and this creates issues with the behavior of the function as $\mathbf{f}$ changes. Given $\mathbf{f} \neq \mathbf{f}'$, $A^{\mathbf{f}} \neq A^{\mathbf{f}'}$, and the preferences over goods contained in one set but not the other may behave strangely. Axioms 4 and 5 restrict how preferences may *change* when the frame moves, but they do not affect goods in $(A^{\mathbf{f}} \cup A^{\mathbf{f}'}) \setminus (A^{\mathbf{f}} \cap A^{\mathbf{f}'})$, because these goods do not see preferences over them change; they see preferences over them being newly revealed. To put order on this process of preference revelation, I use an axiom which imposes the earlier stated desire to treat choices from menus of two goods as complete, transitive, and continuous – *Pairwise Weak Order*. This axiom requires the "pairwise revealed preference" $\succsim^*$, defined by the following:

$$\mathbf{x} \succsim^* \mathbf{y} \Leftrightarrow \mathbf{x} \in C(\{\mathbf{x}, \mathbf{y}\})$$

Note that $\mathbf{x} \succsim^* \mathbf{y} \Leftrightarrow \mathbf{x} \succsim^{\mathbf{f}(\{\mathbf{x},\mathbf{y}\})} \mathbf{y}$, so the completeness of the frame preferences implies the completeness of $\succsim^*$.

Axiom 6 (*Pairwise Weak Order*)**.** $\succsim^*$ is transitive and continuous.

This creates a preference which applies to the whole space, yet also relates the frames to one another, providing more structure to regulate the behavior of the preferences when moving the frame. Along with *Frame Continuity*, *Pairwise Weak Order* is sufficient to guarantee continuity in $\mathbf{f}$, though it is not necessary for continuity in $\mathbf{f}$; this axiom also implies Property 3.

1.3.5 Representation Theorem

Theorem 1 (Compromise/Attraction Representation Theorem)**.** Given a collection of preferences $\{\succsim^{\mathbf{f}}\}_{\mathbf{f} \in \mathbb{R}^2}$, there exists a function $U(\mathbf{x}, \mathbf{f})$ with Properties 1-4 representing it if and only if Axioms 1-6 hold.

Finding a utility function conditional on the frame is trivial; in fact, the first axiom alone guarantees the existence of one. Given a frame $\mathbf{f}$, by *Continuous Weak Order*, there exists $u^{\mathbf{f}}(\mathbf{x})$ which represents the $\mathbf{f}$-preference over $A^{\mathbf{f}}$. Repeat this for each $\mathbf{f} \in \mathbb{R}^2$, and define $U(\mathbf{x}, \mathbf{f}) = u^{\mathbf{f}}(\mathbf{x})$. Of course, this axiom alone does not deliver the other desired properties, namely continuity in $\mathbf{f}$, and therein lies the difficulty.

Proof. The proof first proposes a representation, then establishes some monotonicity and continuity properties of the representation in $\mathbf{f}$. Given a good $\mathbf{x}$ and a frame $\mathbf{f}$ such that $\mathbf{x} \in A^{\mathbf{f}}$, by *Substitutability*, there exists a good $\mathbf{a}(\mathbf{x})$ whose vertical position is f_2 such that $\mathbf{x} \sim^{\mathbf{f}} \mathbf{a}$. By *Simplicity* and transitivity of $\succsim^{\mathbf{f}}$, this $\mathbf{a}$ is unique. Define a function $v_1(\mathbf{x}, \mathbf{f})$ by $\mathbf{x} \sim^{\mathbf{f}} (v_1(\mathbf{x}, \mathbf{f}), f_2)$. $v_1(\mathbf{x}, \mathbf{f})$ is a representation of $\{\succsim^{\mathbf{f}}\}_{\mathbf{f} \in \mathbb{R}^2}$. To see this, consider an $\mathbf{x}'$ such that $\mathbf{x} \succ^{\mathbf{f}} \mathbf{x}'$, $\mathbf{x} \sim^{\mathbf{f}} (v_1(\mathbf{x}, \mathbf{f}), f_2)$ and $\mathbf{x}' \sim^{\mathbf{f}} (v_1(\mathbf{x}', \mathbf{f}), f_2)$, so by transitivity of $\succsim^{\mathbf{f}}$, $(v_1(\mathbf{x}, \mathbf{f}), f_2) \succ^{\mathbf{f}} (v_1(\mathbf{x}', \mathbf{f}), f_2)$. *Simplicity* then implies $v_1(\mathbf{x}, \mathbf{f}) > v_1(\mathbf{x}', \mathbf{f})$.

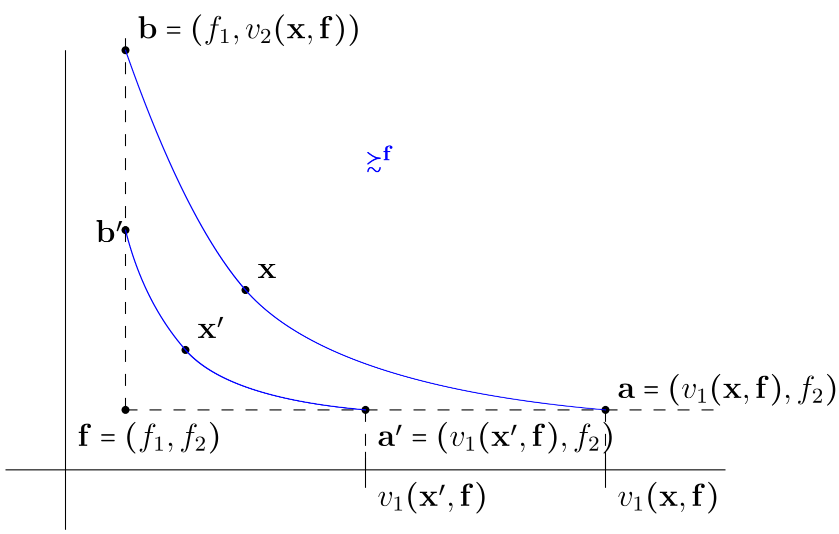

Figure 1·7: Illustration of $v_1(\mathbf{x}, \mathbf{f})$ and $v_2(\mathbf{x}, \mathbf{f})$

By *Simplicity*, $x_i > x_i'$ and $x_{-i} = x_{-i}'$ implies $\mathbf{x} \succ^{\mathbf{f}} \mathbf{x}'$ for each $\mathbf{f}$ such that $\mathbf{x}, \mathbf{x}' \in A^{\mathbf{f}}$. Therefore, v_1 is strictly increasing in x_1 and x_2.

By continuity of $\succsim^{\mathbf{f}}$, v_1 is continuous in x_1 and x_2. To see why, consider $\{\mathbf{x}^n\} \to \mathbf{x}$. Without loss of generality, suppose $\{\mathbf{x}^n\}$ is weakly decreasing in both components.

$v_1(\mathbf{x},\mathbf{f}) \leq v_1(\mathbf{x}^n,\mathbf{f})$, for each n. Therefore,$\inf_n\{v_1(\mathbf{x}^n,\mathbf{f})\} \geq v_1(\mathbf{x},\mathbf{f})$. If the inequality is strict, there exists $y_1 \in (v_1(\mathbf{x},\mathbf{f}), \inf_n\{v_1(\mathbf{x}^n,\mathbf{f})\})$. $\{\mathbf{x}^n\}$ is in the upper contour set of (y_1, f_2) for each n, but $\mathbf{x}$ is not in this upper contour set. This violates *Continuous Weak Order*, and therefore it must be the case that $\inf_n\{v_1(\mathbf{x}^n,\mathbf{f})\} = v_1(\mathbf{x},\mathbf{f})$. This implies $v_1(\mathbf{x}^n,\mathbf{f}) \to v_1(\mathbf{x},\mathbf{f})$, and thus the function is continuous in x_1 and x_2.

Importantly, v_1 is also continuous in f_1.

Lemma 1. Given axioms 1-5, v_1 is continuous in f_1.

Proof. Consider $\{f_1^n\}_{n=1}^{\infty} \to f_1$. Without loss of generality, suppose $\{f_1^n\}_{n=1}^{\infty}$ is increasing[13]. v_1 is decreasing in f_1. By *Compromise/Attraction Monotonicity*, moving the frame right rotates the indifference curves clockwise. Therefore, given $f_1' > f_1$, $\mathbf{f} \equiv (f_1, f_2)$, and $\mathbf{f}' \equiv (f_1', f_2)$, $(v_1(\mathbf{x},\mathbf{f}), f_2) \succ^{\mathbf{f}'} (v_1(\mathbf{x},\mathbf{f}'), f_2)$. This implies $v_1(\mathbf{x},\mathbf{f}) > v_1(\mathbf{x},\mathbf{f}')$, which implies v_1 is decreasing in f_1. Because $\{f_1^n\}_{n=1}^{\infty}$ is increasing, $v_1(\mathbf{x},\mathbf{f}) \leq v_1(\mathbf{x},\mathbf{f}^n)\ \forall\ n \in \mathbb{N}$. As $\{v_1(\mathbf{x},\mathbf{f}^n)\}$ is bounded from below, there exists $\inf_n\{v_1(\mathbf{x},\mathbf{f}^n)\}$, and $v_1(\mathbf{x},\mathbf{f}) \leq \inf_n\{v_1(\mathbf{x},\mathbf{f}^n)\}$.

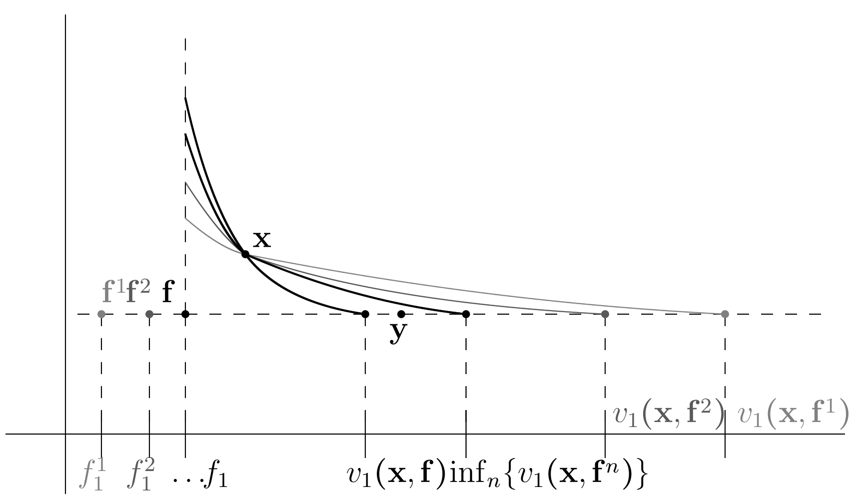

Figure 1·8: Discontinuity of $v_1(\mathbf{x},\mathbf{f})$ Creates a Contradiction

Suppose $v_1(\mathbf{x},\mathbf{f}) < \inf_n\{v_1(\mathbf{x},\mathbf{f}^n)\}$, as illustrated in the above picture. There exists a $\mathbf{y}$ such that $y_2 = f_2$ and $y_1 \in (v_1(\mathbf{x},\mathbf{f}), \inf_n\{v_1(\mathbf{x},\mathbf{f}^n)\})$. $\mathbf{y} \succ^{\mathbf{f}} \mathbf{x}$, but $\mathbf{x} \succ^{(f_1', f_2)} \mathbf{y}$

[13]Because $\{f_1^n\}_{n=1}^{\infty} \to f_1$ is a sequence in $\mathbb{R}$, there exists a monotone subsequence which converges to f_1.

$\forall\ f_1' < f_1$. Furthermore, by *Compromise/Attraction Monotonicity*, $\mathbf{y} \succ^{\mathbf{f}} \mathbf{x}\ \forall\ f_1' > f_1$. Therefore, there is no $\hat{f}_1$ such that $\mathbf{x} \sim^{(\hat{f}_1, f_2)} \mathbf{y}$, and this violates *Frame Continuity*. Thus, $v_1(\mathbf{x}, \mathbf{f}) = \inf\{v_1(\mathbf{x}, \mathbf{f}^n)\}$, it is the limit of $\{v_1(\mathbf{x}, \mathbf{f}^n)\}$, and this implies v_1 is continuous in f_1. □

Symmetric reasoning shows that there is also a function $v_2(\mathbf{x}, \mathbf{f})$ defined by $\mathbf{x} \sim^{\mathbf{f}} (f_1, v_2(\mathbf{x}, \mathbf{f}))$. v_2 is strictly increasing and continuous in x_1 and x_2, and continuous in f_2. So now there are two representations, v_1 and v_2. v_1 has the desired continuity in f_1, and v_2 has the desired continuity in f_2. The next step is to show these functions have the desired continuity in both arguments.

Lemma 2. Given v_1 as defined above, which is continuous in f_1, and v_2 as defined above, which is continuous in f_2, *Pairwise Weak Order* implies v_1 and v_2 are both continuous in both f_1 and f_2.

Proof. Consider the points $(v_1(\mathbf{x}, \mathbf{f}), f_2) \equiv \mathbf{a}$ and $(f_1, v_2(\mathbf{x}, \mathbf{f})) \equiv \mathbf{b}$. Both of these points are $\mathbf{f}$-indifferent to $\mathbf{x}$, which, by transitivity of $\succsim^{\mathbf{f}}$, means $\mathbf{a} \sim^{\mathbf{f}} \mathbf{b}$. Because[14] $\mathbf{f}(\{\mathbf{a}, \mathbf{b}\}) = \mathbf{f}$, this means that $C(\{\mathbf{a}, \mathbf{b}\}) = \{\mathbf{a}, \mathbf{b}\}$, and thus the two points are pairwise indifferent, $\mathbf{a} \sim^* \mathbf{b}$.

$\succsim^*$ is continuous and transitive, and is complete by completeness of the $\mathbf{f}$-preferences. Therefore, there exists a continuous function $u^*(\cdot,\cdot)$ which represents the pairwise preference. Thus,

$$u^*(\mathbf{a}) = u^*(\mathbf{b})$$
$$u^*(v_1(\mathbf{x}, \mathbf{f}), f_2) = u^*(f_1, v_2(\mathbf{x}, \mathbf{f}))$$

$u^*(\cdot,\cdot)$ can be used to demonstrate that v_1 is continuous in f_2. Consider $\{f_2^n\}_{n=1}^{\infty} \to f_2$. v_2 is continuous in f_2, so $\{v_2(\mathbf{x}, \mathbf{f}^n)\}_{n=1}^{\infty} \to v_2(\mathbf{x}, \mathbf{f})$ (where $\mathbf{f}^n \equiv (f_1, f_2^n)$). Therefore, continuity of u^* implies $\{u^*(f_1, v_2(\mathbf{x}, \mathbf{f}^n))\}_{n=1}^{\infty} \to u^*(f_1, v_2(\mathbf{x}, \mathbf{f}))$. For each n, $u^*(v_1(\mathbf{x}, \mathbf{f}^n), f_2^n) = u^*(f_1, v_2(\mathbf{x}, \mathbf{f}^n))$, and $u^*(f_1, v_2(\mathbf{x}, \mathbf{f})) = u^*(v_1(\mathbf{x}, \mathbf{f}), f_2)$. This implies that $\{u^*(v_1(\mathbf{x}, \mathbf{f}^n), f_2^n)\}_{n=1}^{\infty} \to u^*(v_1(\mathbf{x}, \mathbf{f}), f_2)$. Furthermore, by *Simplicity*, $u^*(\cdot,\cdot)$ is strictly increasing in both arguments. Thus, in addition to being continuous in f_1, v_1 is continuous in f_2. The same proof can show that v_2 is continuous in f_1. □

[14] This equality is essential to the proof of this lemma, and will be important in Section 1.3.6, when other frame definitions are discussed.

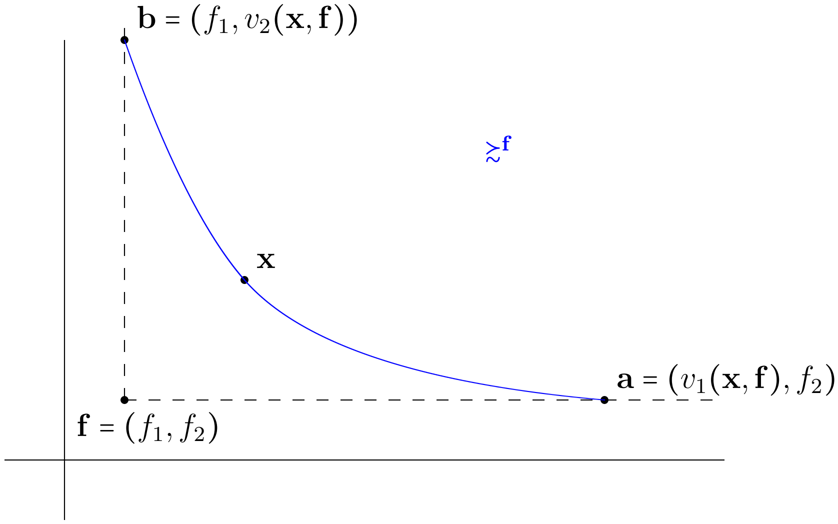

Figure 1·9: Pairwise Continuity Makes v_1 and v_2 continuous in all arguments

There are two functions which satisfy the continuity property required; they each define the function $U(\mathbf{x}, \mathbf{f})$ by

$$\begin{aligned} U(\mathbf{x}, \mathbf{f}) &= u^* \left(v_1(\mathbf{x}, \mathbf{f}), f_2\right) \\ &= u^* \left(f_1, v_2(\mathbf{x}, \mathbf{f})\right) \end{aligned}$$

In other words, the weak order of the pairwise preference "ties together" the v_1 and v_2 functions.

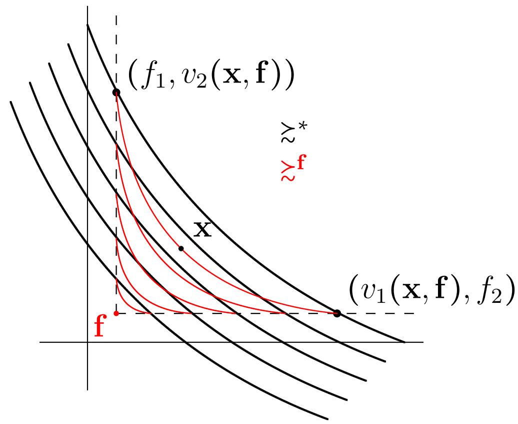

Figure 1·10: Pairwise Continuity Unites v_1 and v_2

The preceding establishes that U satisfies Property 1. U satisfies Property 2 because this property is equivalent to *Compromise/Attraction Monotonicity.* Take a frame $\mathbf{f}$ and a point $\mathbf{x} \in A^{\mathbf{f}}$. Consider the points "left" of $\mathbf{x}$ (i.e., those points with smaller first components) which $\mathbf{x}$ is $\mathbf{f}$-indifferent to. If $\mathbf{f}$ is lowered to some $\mathbf{f}'$ with $f_2' < f_2$ and $f_1' = f_1$, by *Compromise/Attraction Monotonicity*, $\mathbf{x}$ is $\mathbf{f}'$-*preferred* to these points above it. Now consider the points "right" of $\mathbf{x}$ (with a larger first component) which $\mathbf{x}$ is $\mathbf{f}$-indifferent to. $\mathbf{x}$ is $\mathbf{f}'$-*dispreferred* to these points. In other words, the indifference curve passing through $\mathbf{x}$ gets steeper with the move to $\mathbf{f}'$. If instead $\mathbf{f}$ is moved left to some $\mathbf{f}'$ with $f_1' < f_1$ and $f_2' = f_2$, the opposite happens, and the indifference curve gets shallower.

The monotonicity axiom, therefore, can be expressed as a condition on how the slopes of the indifference curves change with respect to changes in $\mathbf{f}$. The slope of the indifference curve is the *marginal rate of attribute substitution (MRAS)*; define the function $MRAS_{\mathbf{x}}(f_1, f_2)$ as the slope of the $\mathbf{f}$ indifference curve at point $\mathbf{x}$[15]. *Compromise/Attraction Monotonicity* is therefore equivalent to:

[15]This does assume that the indifference curve has a slope at point $\mathbf{x}$; however, note that by *Simplicity*, the indifference curves are monotonic. Therefore, they are differentiable almost everywhere. For points in the measure zero set which have no derivative, left and right derivatives can be taken to establish the property.

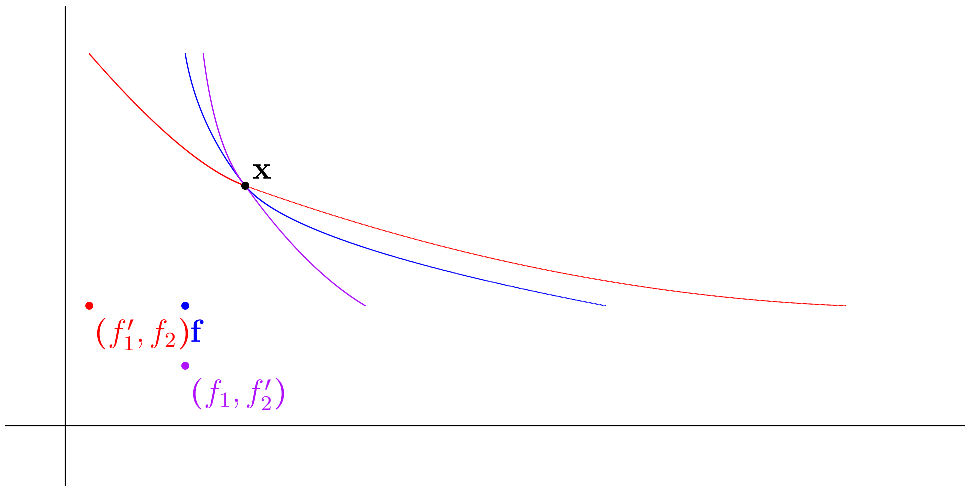

Figure 1·11: The Relationship Between *Compromise/Attraction Monotonicity* and Indifference Curve Slopes

$$\frac{\partial}{\partial f_1} MRAS_{\mathbf{x}}(f_1, f_2) > 0$$
$$\frac{\partial}{\partial f_2} MRAS_{\mathbf{x}}(f_1, f_2) < 0$$

The $MRAS$ is equal to the ratio of the marginal utilities, so $MRAS_{\mathbf{x}}(f_1, f_2) = \frac{U_1(x_1,x_2,f_1,f_2)}{U_2(x_1,x_2,f_1,f_2)}$. Taking these derivatives shows that $U_{13}U_2 - U_1U_{23} > 0 > U_{14}U_2 - U_1U_{24}$, which is equivalent to Property 2[16]:

$$\frac{U_{13}}{U_{23}} > \frac{U_1}{U_2} > \frac{U_{14}}{U_{24}}$$

Transitivity of the pairwise preference is equivalent to Property 3:

$$\left.\begin{array}{r} U(x_1, x_2, x_1, y_2) = U(y_1, y_2, x_1, y_2) \\ U(y_1, y_2, y_1, z_2) = U(z_1, z_2, y_1, z_2) \end{array}\right\} \Rightarrow U(x_1, x_2, x_1, z_2) = U(z_1, z_2, x_1, z_2)$$

Finally, *Substitutability* is equivalent to Property 4. Given a frame $\mathbf{f}$, for each $\mathbf{x} \in A^{\mathbf{f}}$ there exists a $\mathbf{y} \in A^{\mathbf{f}}$ such that

$$U(x_1, x_2, f_1, f_2) = U(f_1, y_2, f_1, f_2) = U(y_1, f_2, f_1, f_2)$$

[16] Assuming U_{23} and U_{24} are not equal to zero; if they are, consider Property 2 to be the non-reduced form.

This then concludes the proof that a representation satisfying properties 1-4 exists *if* Axioms 1-6 are satisfied; it remains to be shown that if the representation exists, it satisfies the axioms.

Given a representation $U(\mathbf{x}, \mathbf{f})$, the usual argument implies *Continuous Weak Order*. Being strictly increasing in x_1 and x_2 implies *Simplicity*. Property 4 implies *Substitutability*. Property 3 implies Pairwise Transitivity, and Property 2 implies *Compromise/Attraction Monotonicity*. These are all trivial. The remaining axioms (*Frame Continuity* and Pairwise Continuity) require more sophisticated arguments.

Consider $\{(\mathbf{y}^n)\}_{n=1}^{\infty} \to (\mathbf{y})$, where $\mathbf{x} \sim^* (\mathbf{y}^n)$ for each n. Pairwise Continuity is implied if $\mathbf{x} \sim^* (y_1, y_2)$. By the continuity of U, $U(x_1, x_2, x_1, y_2^n) \to U(x_1, x_2, x_1, y_2)$ (Assuming, without loss of generality, that $\mathbf{x} \succ_2 \mathbf{y}$). Also by continuity, $U(y_1^n, y_2^n, x_1, y_2^n) \to U(y_1, y_2, x_1, y_2)$[17]. However, $U(x_1, x_2, x_1, y_2^n) = U(y_1^n, y_2^n, x_1, y_2^n)\ \forall\ n$. This implies that $U(x_1, x_2, x_1, y_2) = U(y_1, y_2, x_1, y_2)$, as desired.

As for *Frame Continuity*, consider, $\mathbf{x}, \mathbf{y}, \mathbf{f}$ such that $\mathbf{x} \succ^{\mathbf{f}} \mathbf{y}$, and f_i' such that $\mathbf{y} \succ^{(f_{-i}, f_i')} \mathbf{x}$. Suppose there does not exist an f_i'' such that $\mathbf{x} \sim^{(f_{-i}, f_i'')} \mathbf{y}$. In other words, while there is no f_i which makes $\mathbf{x} \sim^{\mathbf{f}} \mathbf{y}$, there are f_i's which make it both preferred and dispreferred. So as f_i is lowered and the indifference curve rotates through $\mathbf{x}$ as per *Compromise/Attraction Monotonicity*, there is a gap in the area covered by the indifference curves, and $\mathbf{y}$ resides in this gap.

This case is straightforwardly ruled out by the continuity of U. If such a gap

[17]This does require U to be *jointly* continuous in all arguments; it can be seen that this is implied by the continuity in each individual argument as follows: to show $\{(x_1^n, x_2^n, f_1^n, f_2^n)\}_{n=1}^{\infty} \to (x_1, x_2, f_1, f_2) \Rightarrow \{U(x_1^n, x_2^n, f_1^n, f_2^n)\}_{n=1}^{\infty} \to U(x_1, x_2, f_1, f_2)$, break up the convergent sequence into $\{x_1^n\}_{n=1}^{\infty} \to x_1$, $\{x_2^n\}_{n=1}^{\infty} \to x_2$, $\{f_1^n\}_{n=1}^{\infty} \to f_1$, and $\{f_2^n\}_{n=1}^{\infty} \to f_2$. With these constituent parts, we can see:

$$\begin{aligned}
&\exists\ N_1 \text{ s.t.} \quad \forall\ n_1 > N_1, \quad |w(x_1, x_2, f_1, f_2^{n_1}) - w(x_1, x_2, f_1, f_2)| < \frac{\varepsilon}{4} \\
&\exists\ N_2 \text{ s.t.} \quad \forall\ n_2 > N_2, \quad |w(x_1, x_2, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1, f_2^{n_1})| < \frac{\varepsilon}{4} \\
&\exists\ N_3 \text{ s.t.} \quad \forall\ n_3 > N_3, \quad |w(x_1, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1^{n_2}, f_2^{n_1})| < \frac{\varepsilon}{4} \\
&\exists\ N_4 \text{ s.t.} \quad \forall\ n_4 > N_4, \quad |w(x_1^{n_4}, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2^{n_3}, f_1^{n_2}, f_1^{n_1})| < \frac{\varepsilon}{4}
\end{aligned}$$

By the Δ-inequality, $|w(x_1^{n_4}, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1, f_2)| \leq |w(x_1^{n_4}, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2^{n_3}, f_1^{n_2}, f_2^{n_1})| + |w(x_1, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1^{n_2}, f_2^{n_1})| + |w(x_1, x_2, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1, f_2^{n_1})| + |w(x_1, x_2, f_1, f_2^{n_1}) - w(x_1, x_2, f_1, f_2)|$. If we choose $n_i > \max\{N_1, N_2, N_3, N_4\}\ \forall\ i \in \{1, 2, 3, 4\}$, then the RHS of the above inequality is $< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon$.

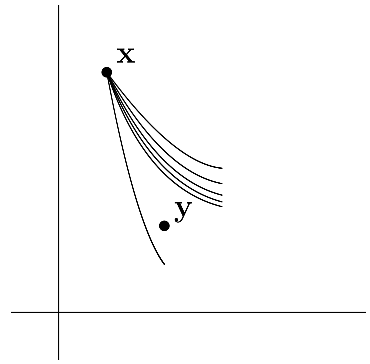

Figure 1·12: Absence of *Frame Continuity* Creates a Contradiction

exists, there exists $\{f_i^n\}_{n=1}^{\infty} \to \hat{f}_i$ such the following two statements hold:

$$U(x_1, x_2, f_{-i}, f_i^n) > U(y_1, y_2, f_{-i}, f_i^n) \ \forall \ n \tag{1.3.4}$$

$$U(x_1, x_2, f_{-i}, \hat{f}_i) < U(y_1, y_2, f_{-i}, \hat{f}_i) \tag{1.3.5}$$

However, by the continuity of U,

$$\lim_{n\to\infty} U(x_1, x_2, f_{-i}, f_i^n) = U(x_1, x_2, f_{-i}, \hat{f}_i) \tag{1.3.6}$$

$$\lim_{n\to\infty} U(y_1, y_2, f_{-i}, f_i^n) = U(y_1, y_2, f_{-i}, \hat{f}_i) \tag{1.3.7}$$

This contradicts (1.3.4) and (1.3.5), and therefore such a gap cannot exist.

□

1.3.6 Alternate Frame Definitions

Up to this point, I have used a very specific definition of the frame. However, my representation theorem is robust to a variety of approaches to defining the frame. This is useful because there are examples where the most appropriate frame definition is not obvious, such as Figure 1·13.

z is dominated on both attributes by both **x** and **y**. As the attraction effect is

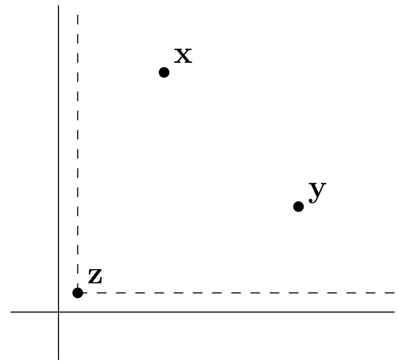

Figure 1·13: Three Good Menu Motivating Alternate Frame Definition

normally described in an example where the third good is only dominated by *one* other good, and thereby induces the choice of that good, it is unclear what the effect of **z** should be in this situation. The definition of the frame used up until this point suggests that it may induce a different choice, though which choice is unclear as it affects both goods. This seems plausible. One could also argue that a decision maker would completely ignore **z**, and only focus on **x** and **y**, and therefore it should have no impact. This also seems plausible. *Frame Preferences* are robust to the following definition of frame that adopts this latter interpretation of behavior:

$$\text{Define } \hat{S} \equiv \{x \in S \mid \exists\, y \in S,\ i \in \{1,2\} \text{ such that } x \succ_i y\}$$
$$\mathbf{f}(S) \equiv \left(\min_{\mathbf{x}\in\hat{S}} x_1, \min_{\mathbf{x}\in\hat{S}} x_2\right)$$

This two-step frame definition first winnows the menu by removing all points such as **z** in the picture above, which are dominated on both attributes by all other goods. It then applies the original frame definition to the winnowed menu. A collection of revealed frame preferences can be constructed using this definition of frame, and the representation theorem will still hold for this new collection of frame preferences.

In fact, any frame definition and associated collection of revealed frame preferences which satisfy the following weak condition will generate a collection of revealed frame preferences for which the representation theorem holds:

Condition 1 (*Pairwise Edge Consistency*)**.** Given $(v_1, f_2), (f_1, v_2)$ such that $(v_1, f_2) \sim^{\mathbf{f}} (f_1, v_2)$, $\mathbf{f}\left(\{(v_1, f_2), (f_1, v_2)\}\right) = \mathbf{f}$.

This works because the proof of the representation theorem only relies on the definition of the frame in one place. After having established the representations v_1 and v_2, it is noted that, by transitivity of $\succsim^{\mathbf{f}}$, $\mathbf{a} \equiv (v_1(\mathbf{x}, \mathbf{f}), f_2) \sim^{\mathbf{f}} (f_1, v_2(\mathbf{x}, \mathbf{f})) \equiv \mathbf{b}$. Because[18] $\mathbf{f}(\{\mathbf{a}, \mathbf{b}\}) = \mathbf{f}$, this implies $\mathbf{a} \sim^* \mathbf{b}$. That is all the representation theorem requires of the frame definition; that the menu consisting solely of the two $\mathbf{f}$-indifferent goods used to identify v_1 and v_2 also have the frame $\mathbf{f}$, thus guaranteeing their pairwise indifference.

Indeed, there are many frame definitions which would generate a collection of preferences with a representation satisfying the given properties. However, many of these frame definitions would not combine with *Compromise/Attraction Monotonicity* to produce an outcome consistent with the compromise and attraction effects. For example, defining the frame as the maximum along each attribute would allow for the compromise effect, but not the attraction effect (adding $\mathbf{z}$ would shift the frame, but adding $\mathbf{z}'$ would not).

However, reference dependence abounds, and there may be other effects best illuminated through other frame definitions. Combining a frame definition and associated collection of frame preferences which satisfy *Pairwise Edge Consistency* with *Pairwise Weak Order* can be used as a basis for creating a utility function which is continuous in the frame. Specifically, if the behavior of the preferences is such that there exist representations v_i along the edge of the frame as defined in the proof, and each v_i is continuous in f_i, then *Pairwise Edge Consistency* and *Pairwise Weak Order* make

[18]By construction, $v_1(\mathbf{x}) > f_1$, and $v_2(\mathbf{x}) > f_2$.

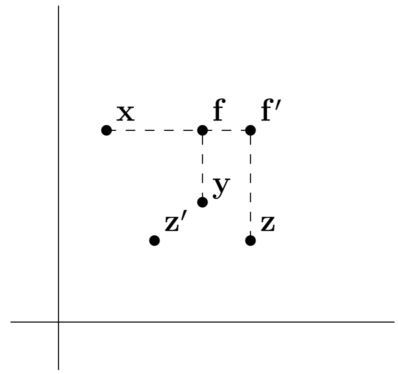

Figure 1·14: Failure of Defining Frame by Maximum to Produce Attraction Effect

those representation functions continuous in all components of $\mathbf{f}$. This is a potentially useful construct for further study of reference dependence.

1.4 Conclusion

Though the compromise and attraction effects have long been well established experimentally, there is little available in the way of systematic representations. This paper offers a fully axiomatized model which incorporates both effects, by embracing the notion of frame preferences and using them to create a succinct expression of the effects which has a natural mathematical interpretation, summarized by the *Compromise/Attraction Monotonicity* axiom and the *Compromise/Attraction Rotation* property.

Through its regularity properties, this model is very convenient for applications. It is natural to consider compromise and attraction effects in an industrial organization setting, for example in the case of a multi-product monopolist, or a single product oligopoly game. Ok et al. (2011) do the former, using their model which features the attraction effect but not the compromise effect. The latter situation remains unexplored. In both cases, this model would be ideal for exploring optimal strategies in the presence of both effects.

The compromise and attraction effects can also be applied to the political realm. Pan et al. (1995) show evidence of the attraction effect in voting results from the 1992 U.S. presidential election and the 1994 Illinois Democratic gubernatorial primary. Relatedly, it is a commonly observed phenomenon that many elections feature extremist candidates with no possibility of winning. While no other papers provide explanations for this result, Poterack and Solow (2015) use Frame Preferences to do so.

In addition, this paper makes behavioral predictions which can be tested experimentally. In particular, while papers have studied both compromise and attraction effects, there has not been an attempt to construct a single experiment which can show either effect in the same setting. Such an experiment could be used to estimate the relative magnitudes of the effects, which this paper predicts would be the same. Also, the effect of adding goods which do not change the frame should be explored; i.e., it should be tested whether the addition of non-extreme goods can induce either effect.

Finally, as noted the paper draws a relationship between *Pairwise Weak Order* and a representation which is continuous in the frame. As reference dependence is common, and as constructing a utility which is continuous in the reference is desirable, this is a potentially useful basis for future papers looking to construct convenient reference dependent representations.

Chapter 2

Extremist Politics and the Preference for Compromise*

2.1 Introduction

Violations of the Weak Axiom of Revealed Preference (or Independence of Irrelevant Alternatives) have been extensively documented in choice from menus across many domains. Two particular violations, the "compromise effect" and the "attraction effect," seem to be particularly robust features of behavior. The compromise effect refers to a tendency of decision-makers to avoid choices that are "extreme" in some feature in the choice set. The attraction effect can be described as a tendency for decision-makers to select an option which strictly dominates some other option in the choice set if no other alternatives satisfy the same dominance relation.

In this paper, we extend the canonical citizen-candidate model to a multi-dimensional policy space and consider how the presence of voters who are subject to compromise and attraction effects change incentives for candidate entry. Despite having observed these effects in many different contexts of choice from menus, the literature has largely ignored how the presence of these behavioral effects change incentives for optimal menu construction. In particular, while Pan, et al (1995) documented the presence of these effects in political contests, the literature has yet to incorporate voter preferences generating compromise or attraction effects in models of political competition.

*This chapter of the dissertation is based on joint work with Benjamin L. Solow.

We show that incorporating these preferences generates novel incentives for candidate behavior and provides a strategic incentive for extremist candidates to run for office in equilibrium.

We build a tractable model of endogenous entry into plurality elections where voters act sincerely, but have preferences subject to the compromise and attraction effects. We model the compromise and attraction effects via "frames of reference" as in Tserenjigmid (2015) and Poterack (2015). Voters have horizontally differentiated single-peaked preferences along two dimensions of policy, but different menus of candidates on the ballot can generate different frames of reference and thus change how voters evaluate the policy platforms of the candidates.

Elections are of particular interest in the study of supply-side responses to the compromise and attraction effects. Candidates can be thought of as single-product oligopolists competing for market share, but without access to price as an instrument. Differing from competitive product markets, however, there is a clear menu facing all consumers in the market which is endogenously determined by strategic behavior. In many contexts in traditional product market competition, understanding what menu a consumer actually observes is difficult, yet integral to understanding these effects. Thus, our model should be understood to provide some insight into the forces introduced to oligopolistic product markets by these effects while clearly leaving scope for further work.

This paper makes three main contributions. First, we provide a tractable model of spatial competition in elections with multiple dimensions of policy space. Failures of canonical models to generalize from single to multi-dimensional policy spaces has plagued the study of electoral politics. In particular, Plott (1967), Kramer (1973), McKelvey (1976) and others demonstrated that pure strategy Nash equilibria generally fail to exist when candidates compete over policy spaces with more than one ideological dimension. Solutions to this problem have included introducing uncertainty to voter behavior in order to smooth payoff functions (Roemer 2004), eliminating either the spatial structure or single-peakedness assumptions on voter utility functions (Besley and Coate 1997), or adding candidate-specific valence effects (Ansolabehere and Snyder 2000). We require neither type of assumption and instead obtain our results by restricting the degree of commitment available to candidates. This is a natural assumption that reflects realities of political competition at high levels of elected office. First, candidates do not, in reality, have commitment technologies available to them. Second, voters likely have well-formed beliefs about the policies a (for example, presidential) candidate would implement based on an observed history of policymaking.[1]

Second, we show general results on the incentives generated by the compromise and attraction effects in spatial models. We show, independent of the chosen functional form of the representation, asymmetric effects of extremist entrants on moderate candidates. For symmetric distributions of voter preferences, however, any particular extremist entrant has an "equal and opposite" extremist who has exactly countervailing effects on voter preferences over the moderate candidates. For skewed distributions of voter preferences, this equal and opposite extremist may not exist. Therefore, we conclude that distance from the mean voter's preferred policy may be relatively more important than distance from the median voter's preferred policy in multicandidate elections.

Third, we implement a particular functional form of these effects characterized by Tserenjigmid (2015) and Poterack (2015) and use that form to study Nash equilibria of

[1] One can think of our game as being a stage of a dynamic game where politicians invest in being associated with a policy platform early in their career by voting for it when in office. Later, voters would have more precise views about the candidates' policy preferences. So long as there is still a payoff to being associated with this policy after leaving office (e.g. legacy concerns, transitioning to private lobbying work, etc.), politicians would not deviate in this stage game either.

an entry game. We show that there exist linear equilibria where extremists enter and obtain office with probability zero. By entering, however, they positively frame their preferred moderate candidate, and thus shift the expected policy outcome closer to their ideal point. This is a novel result in models of spatial competition. All previous models which generate sure loser equilibria either require them to be centrist with respect to the competitive candidates in a "squeezing" equilibrium (Osborne and Slivinski 1996, Solow 2015) or require violations of single-peaked preferences (Besley and Coate 1997).

2.2 Related Literature

Multidimensional spatial competition has been of great interest historically, especially in the domain of supply-side behavior in elections. Unfortunately, early work by Plott (1967), McKelvey (1976) and others demonstrated that in Hotelling-Downs models of multidimensional competition, pure strategy Nash equilibria generally do not exist. Despite this, multidimensional competition seems important. Ahler and Broockman (2016) argues that mapping voters' preferences to a single dimension can mischaracterize as moderate voters who hold immoderate views on a number of issues, but are less correlated in their preferences than parties are.

Theorists have employed a number of strategies to attempt to address the problems with multidimensional Hotelling-Downs. Ansolabehere and Snyder (2000) introduce valence, a vertical differentiation component, to the model and show that valence issues restore equilibria. In all equilibria, however, the candidate with a greater valence score wins the election with probability one. Mckelvey (1986) utilizes a different solution technique for the game and shows that the uncovered set contains equilibrium behavior under several different institutional structures. Roemer (2004) studies the Hotelling-Downs model where voters behave probabilistically, which smooths the discontinuities in candidate payoff functions which cause Nash equilibria not to exist.

Our results here are complementary. Rather than introduce noise to voter preferences or change the mode of differentiation, we restrict candidates' ability to commit to different policy platforms. As in Osborne and Slivinski (1996), candidates are unable to credibly commit to a policy platform different than their most preferred policy. We show that this lack of commitment technology restores existence of equilibria and that some of these equilibria address undesirable features of the set of equilibria in other models. For example, all candidates who choose to enter despite losing with certainty in Osborne and Slivinski (1996) must be centrist. This is no longer the case in a multidimensional issue space. Additionally, the multidimensional domain of competition allows us to study the attraction and compromise effects.

The attraction effect was first demonstrated experimentally in the early 1980's by Huber et al. (1982), and the compromise effect was demonstrated first by Itamar Simonson (1989), whose paper also provided support for the attraction effect. These papers are strictly concerned with observing the effects; neither of them attempt to construct a representation incorporating these effects. Some effort to do so is undertaken by Simonson and Tversky (1993), whose model (unlike ours) depends on context created by all elements in a menu, not just the worst along each attribute. Kivetz et al. (2004) analyze the Simonson and Tversky model, and two others, in terms of which best fit available experimental data.

Recently, there has been renewed interest in applying models of the attraction and compromise effects to economic questions. In particular, Ok et al. (2015), utilize their representation of the attraction effect to derive the optimal menu of vertically differentiated products offered by a multiproduct monopolist in the presence of consumers who exhibit the attraction effect. Our contribution is complementary. We provide the first, to our knowledge, study of horizontal differentiation in the presence of the compromise effect. In addition, our framework is competitive rather than monopolistic, although we do not study the implications of these effects for pricing.

The most closely related paper to our application here is Pan, et al. (1995). Pan, et al. argue that experimental subjects exhibit the attraction effect when given menus of political candidates who have been scored on various issues. The subjects were randomly assigned to menus where different candidates had an asymmetric dominance relationship with a third candidate. Despite subjects' previous familiarity with the candidates (e.g. presidential candidates), Pan, et al. recover evidence of the attraction effect in voter choice.

2.3 Model

We begin with the standard continuous citizen-candidate model of Osborne and Slivinski (1996). In any election, the electorate is comprised of a unit mass of citizens $\mathcal{I}$ with single-peaked preferences over policy. Policies are represented by the double (e, s); these two dimensions can be thought of representing, for example, economic and social policies. The ideal point for citizen i is denoted (e^i, s^i), and the ideal points of the electorate are distributed over $\mathbb{R}^2$ according to an arbitrary distribution function F. We assume F is continuous and strictly increasing in both arguments. In Section 2.5 we will restrict F to be bounded and have support $[0,1] \times [0,1]$.

We also assume there exists a unit mass of potential candidates $\mathcal{J}$ with single-peaked preferences over policy. Potential candidates are also distributed over $\mathbb{R}^2$ according to continuous, strictly increasing distribution function G, where support(G) = support(F).[2] The action set for potential candidate j is denoted $\mathcal{A}_j = \{E,\ N\}$ where E represents entering the race, and N represents not entering. Types are common

[2]We define the set of citizens and candidates separately in order to facilitate the analysis of Section 2.5. Our results from Section 2.4 would be identical if candidates were also allowed to vote. Our restriction on the supports is also an analytical convenience and not required to retain the qualitative properties of our results.

knowledge, and no commitment technology is available. Since there is a unit mass of citizens, no single citizen may be pivotable between candidates, and therefore all citizens vote sincerely. If a potential candidate chooses to enter the race, we call her a candidate.

Citizens are policy motivated and have no preference over the identity of any candidate (i.e. there are no valence effects). Potential candidates are also policy motivated. In addition, those candidates who choose to stand for office incur a utility cost c of running for office.[3] Whichever candidates wins the election obtains office-related benefits b (e.g. "ego-rents" as in Rogoff 1990). We assume policy motivations take the form of a linear cost function in the difference between the implemented policy and a citizen or candidate's ideal policy. Therefore, if the winner of the election has ideal policy (e^w, s^w), citizens and all potential candidates who choose N and have type $\tau_j = (e^j, s^j)$, obtain a payoff of

$$\pi_j(N, \sigma_{-j}; \tau_j = (e^j, s^j)) = -((e^j - e^w)^2 + (s^j - s^w)^2)^{\frac{1}{2}}$$

A candidate who chooses to enter the race obtains an additional payoff b in the event of winning the election, but pays cost c regardless of the outcome. Therefore, for a candidate with ideal policy (e^j, s^j), her payoff of choosing E is

$$\pi_j(E, \sigma_{-j}; \tau_j = (e^j, s^j)) = \begin{cases} b - c & \text{if wins outright} \\ -((e^j - e^w)^2 + (s^j - s^w)^2)^{\frac{1}{2}} - c & \text{if loses outright.} \end{cases}$$

Henceforth, we use $d(x, y)$ to denote the Euclidean distance (policy relevant portion

[3]The cost of running for office is better understood as the *net cost* of running for office absent parameters modeled here. If candidates obtained, for example, dynamic career benefits from running, that would appear in this model as a lower value of c. The assumption we make of equal cost of running for office, therefore, excludes cases such as term-limited candidates or a mapping from vote shares to post-electoral influence.

of payoffs) between x and y. Each potential candidate obtains a payoff of $-\infty$ if no one chooses to enter. The office-related portion of the payoff function, b, therefore represents the magnitude of the incentives to run from obtaining office relative to the incentives to run from the ability to affect policy.

The game is a simultaneous move entry game. Thus, a strategy is a mapping $\sigma_j : \tau_j \to \mathcal{A}_j$.. As is typical in citizen-candidate models, we focus on characterizing pure strategy Nash equilibria by the number of candidates who choose to enter (e.g. the set of equilibria where two candidates choose to enter the race). An equilibrium is given by $\pi_j(\sigma_j, \sigma_{-j}) \geq \pi_j(\sigma'_j, \sigma_{-j})$ for all j and σ'_j. Let the set of candidates be given by $\mathcal{K} = \{j \in \mathcal{J} : \sigma_j = E\}$ and let the vector of types for the set of candidates who have chosen to enter be denoted $\boldsymbol{\tau}_K$. Denote by $v_j(\boldsymbol{\tau}_K)$ the vote share of candidate j, i.e. the measure of voters who prefer j to all other candidates. We study plurality elections. Therefore, the set of victors is given by $W = \{j : v_j \geq v_k \forall j, k \in K\}$.

2.4 Multidimensional Citizen-Candidate

As is traditional in citizen-candidate models, we characterize sets of pure strategy Nash equilibria by the number of candidates who choose to enter the race. The relative magnitudes of the set of two candidate equilibria and the sets of multi-party (three or more) candidate equilibria are of particular interest. Duverger's (1954) Law, the statement that plurality rule generates two party systems, seems to hold empirically. Nevertheless, three candidate equilibria do exist in plurality elections in the unidimensional citizen-candidate model. Osborne and Slivinski (1996) show a form of Duverger's Law: for all distributions of voter preferences F, the set of parameters b and c that support two candidate equilibria under the runoff rule are a subset of those that support two candidate equilibria under plurality rule. While we do not consider the runoff rule here, our analysis provides the first set of results necessary to compare the two rules in the citizen-candidate model under multidimensional competition.

As is the case in all citizen-candidate models, we face a large multiplicity of equilibria. Indeed, the move to multiple dimensions of competition makes this problem worse, as there are more sets of candidate locations that generate vote shares that may occur in equilibrium. The first step, therefore, is to attempt to limit the potential positions that candidates can take in equilibrium. In a single dimension, candidates who lose with certainty cannot be extreme amongst the set of entrants. This result is due to the fact that a candidate who is losing with certainty, but extreme among the set of entrants, either has no effect on the policy implemented or helps her least preferred alternative candidate win the election. This no longer holds in a two dimensional model, but the intuition of the result remains.

Lemma 1. There does not exist an equilibrium in which a candidate x loses with certainty if $\forall\ i$ such that $\arg\max_{z\in K} u^i(e^z, s^z) = x$, $\arg\max_{z\in K/\{x\}} u^i(e^z, s^z) = y$ for some $y \in K/\{x\}$.

Lemma 1 formalizes the underlying logic of Lemma 1 in Osborne and Slivinski (1996) and extends it to multiple dimensions. In this multi-dimensional framework we can see clearly that the restrictions on positions of sure losers in equilibrium is not really about the relative extremity of candidates, but rather about *agreement* of their voters on the next-best alternative.

This is particularly relevant to models of strategic voting with exogenous candidates which often study "divided majority" cases (e.g. Myerson and Weber 1993). In a divided majority setting with three candidates, a majority of voters prefer either of candidates A or B to candidate C, but disagree on their ranking of A and B. Our result suggests that if candidates are policy motivated as well as office motivated, and agree with their supporters on policy, the assumption of both candidates A and B choosing to participate in the election is not innocuous.

We wish to characterize the set of multi-candidate equilibria. In order to do so, we must first characterize vote shares for entry by arbitrary sets of candidates. As in McKelvey (1986) and Ansolabehere and Snyder (2000), our voters' preferences here are Euclidean. Thus, the set of voters who are indifferent between two candidates x and y is defined by the perpendicular bisector of the vector $\tau_x - \tau_y$. Let M denote the set of all median lines, i.e. all lines that divide $F(.)$ into two regions of equal mass. Borrowing a term from McKelvey (1986), let the "yolk" of the distribution be defined as the smallest ball in $\mathbb{R}^2$ which intersects all median lines. When specializing the distribution in later sections, we will return to the importance of the yolk for existence of different equilibria.

In order to characterize the set of two candidate equilibria, we must know the vote shares of all configurations with at least three candidates. Suppose there exist three candidates, x, y, and z, with $(e^x, s^x) = (\frac{1}{2} + a, \frac{1}{2} + b)$, $(e^y, s^y) = (\frac{1}{2} + c, \frac{1}{2} + d)$, and $(e^z, s^z) = (\frac{1}{2} + f, \frac{1}{2} + g)$. Without loss of generality, let $b < d$. The set of voters who are indifferent between candidates x and y is given by $s^i = \frac{a-c}{d-b}e^i + \frac{1+d+b}{2} - \frac{a-c}{d-b}\left(\frac{1+a+c}{2}\right)$, and therefore the set of voters who prefer x to y are all voters with $s^i \leq \frac{a-c}{d-b}e^i + \frac{1+d+b}{2} - \frac{a-c}{d-b}\left(\frac{1+a+c}{2}\right)$. The set of voters who are indifferent between x and z is given by $s^i = \frac{a-f}{g-b}e^i + \frac{1+d+g}{2} - \frac{a-f}{g-b}\left(\frac{1+a+f}{2}\right)$. Finally, the set of voters who are indifferent between y and z are given by $s^i = \frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}\left(\frac{1+c+f}{2}\right)$.

Candidate i's voters are given by the set of voters that prefer her to every other candidate. If $g > d$,

$$\begin{aligned}
\nu_x = \{i : s^i \leq \min\{&\frac{a-f}{g-b}e^i + \frac{1+b+g}{2} - \frac{a-f}{g-b}(\frac{1+a+f}{2}),\\
&\frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}(\frac{1+c+f}{2})\}\}\\
\nu_y = \{i : &\frac{a-c}{d-b}e^i + \frac{1+d+b}{2} - \frac{a-c}{d-b}(\frac{1+a+e}{2}) < s^i \leq\\
&\frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}(\frac{1+c+f}{2})\}\\
\nu_z = \{i : s^i > \max\{&\frac{a-f}{g-b}e^i + \frac{1+b+g}{2} - \frac{a-f}{g-b}(\frac{1+a+f}{2}),\\
&\frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}(\frac{1+c+f}{2})\}\}
\end{aligned}$$

Similarly, if $g < b$,

$$\begin{aligned}
\nu_x = \{i : &\frac{a-f}{g-b}e^i + \frac{1+b+g}{2} - \frac{a-f}{g-b}(\frac{1+a+f}{2}) < s^i \leq\\
&\frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}(\frac{1+c+f}{2})\}\}\\
\nu_y = \{i : s^i > \max\{&\frac{a-c}{d-b}e^i + \frac{1+d+b}{2} - \frac{a-c}{d-b}(\frac{1+a+e}{2}),\\
&\frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}(\frac{1+c+f}{2})\}\}\\
\nu_z = \{i : s^i \leq \min\{&\frac{a-f}{g-b}e^i + \frac{1+b+g}{2} - \frac{a-f}{g-b}(\frac{1+a+f}{2}),\\
&\frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}(\frac{1+c+f}{2})\}\}
\end{aligned}$$

Finally, if $b < g < d$,

$$\nu_x = \{i : s^i \leq \min\{\frac{a-f}{g-b}e^i + \frac{1+b+g}{2} - \frac{a-f}{g-b}(\frac{1+a+f}{2}),$$
$$\frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}(\frac{1+c+f}{2})\}\}$$
$$\nu_y = \{i : s^i > \max\{\frac{a-c}{d-b}e^i + \frac{1+d+b}{2} - \frac{a-c}{d-b}(\frac{1+a+e}{2}),$$
$$\frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}(\frac{1+c+f}{2})\}\}$$
$$\nu_z = \{i : \frac{a-f}{g-b}e^i + \frac{1+b+g}{2} - \frac{a-f}{g-b}(\frac{1+a+f}{2}) < s^i \leq$$
$$\frac{c-f}{g-d}e^i + \frac{1+d+g}{2} - \frac{c-f}{g-d}(\frac{1+c+f}{2})\}$$

Therefore, vote shares v_x, v_y, and v_z are given simply by integrating $F(.)$ over the subsets of policy space that prefer each candidate.

Consider two candidates, x and y, with ideal policies (e^x, s^x) and (e^y, s^y). Without loss of generality, assume $e^x < e^y$ and $s^x < s^y$. We call an entrant z with ideal policy (e^z, s^z) *extreme with respect to party competition* if either $\tau_z << \tau_x$ or $\tau_z >> \tau_y$. Denote by $d^*(F)$ the critical value of $d(x,y)$ such that if $d(x,y) > d^*(F)$, there exists an entrant who is not extreme with respect to party competition who can enter and win the election and if $d(x,y) < d^*(F)$ no such entrant exists.[4]

Proposition 1. 1. In any two candidate equilibrium, $d(x,y) < d^*(F)$ and $s^i = \frac{s^x+s^y}{2} + \frac{e^x-e^y}{s^y-s^x}\left(e^i - \frac{e^x+e^y}{2}\right)$ is a median line.

2. There exists a two candidate equilibrium with candidates x and y located at (e^x, s^x) and (e^y, s^y) if and only if $v_x = v_y = \frac{1}{2}$, $d(x,y) > 0$, $b \geq 2c - d(x,y)$, for all i such that $v_x(x,y,i) > v_y(x,y,i)$, $c \geq \frac{1}{2}\left[d(y,i) - d(x,i)\right]$ and for all i such that $v_y(x,y,i) > v_x(x,y,i)$, $c \geq \frac{1}{2}\left[d(x,i) - d(y,i)\right]$, and either $d(x,y) < d^*(F)$ or $d(x,y) = d^*(F)$ and for all i who cause $d^*(F)$ to bind, $b \leq 3c - \frac{1}{2}(d(x,i) + d(y,i))$.

Two candidate equilibria share similar features with the unidimensional model. In particular, the line dividing the two groups of voters must be an element of M.

[4] Note that given continuity of $F(.)$, $d^*(F) > 0$.

In a single dimension, the candidates must be equidistant from the median voter. As opposed to Ansolabehere and Snyder (2000), there cannot exist a two candidate equilibrium where a candidate wins with certainty. This is not the case with more than two candidates.

Equilibria with three candidates do exist in plurality elections as well. Notably, the move to multiple dimensions generates three candidate equilibria where a single candidate wins the election with certainty. This equilibrium is very similar to an example given in Besley and Coate (1996). Our contribution is to show that their example is a much more robust potential outcome and does not require the violations of single-peakedness and narrow assumptions on preferences they use. Furthermore, we identify a constant and undesirable feature of this class of equilibria: all sure winners in three candidate elections must be Condorcet losers among the set of entrants.

Proposition 2. Let $K = \{x, L_1, L_2\}$. There exists a three candidate equilibrium with a sure winner, x, and sure losers, L_1, L_2, if and only if:

1. $v_x > \max\{v_{L_1}; v_{L_2}\}$.
2. $\forall$ z such that $v_z = \max_{k \in K \cup \{z\}} v_k$, $b \leq c - d(x, z)$.
3. $\forall$ z such that $v_{L_i} = \max_{k \in K \cup \{z\}}$, $c \geq d(x, z) - d(L_i, z)$.
4. $c \leq \min\{d(L_1, L_2) - d(L_1, x); d(L_1, L_2) - d(L_2, x)\}$.
5. x is a Condorcet loser, i.e. $v_x(x, L_1) < v_{L_1}(x, L_1)$ and $v_x(x, L_2) < v_{L_2}(x, L_2)$.
6. $b \geq c - d(x, L^*)$, where $L^* = \arg\max_{L_i} v_{L_i}(L_1, L_2)$.

Additionally, there exist equilibria where a single candidate loses with certainty. Here a crucial distinction with respect to unidimensional competition arises. As Lemma 1 hints, being extreme with respect to the center of the distribution actually has no bearing on whether a candidate may find it optimal to enter as a sure loser. Instead, what matters is that they do not cannibalize vote share exclusively from their preferred alternative candidate.

Proposition 3. There exists an equilibrium with three candidates, x, y, and L, where only candidate L loses with certainty if and only if

1. $v_x(x, y, L) = v_y(x, y, L) > v_L(x, y, L)$.

2. $d(x, L) \neq d(y, L)$.

3. $b \geq 2c - d(x, y)$.

4. If $d(x, L) < d(y, L)$, $v_x(x, y) < v_y(x, y)$ and $d(y, L) - d(x, L) \geq c$.

5. If $d(x, L) > d(y, L)$, $v_x(x, y) > v_y(x, y)$ and $d(x, L) - d(y, L) \geq c$.

6. $f(v_L)$ is not symmetric across the line $s^i = \frac{s^x + s^y}{2} + \frac{e^x - e^y}{s^y - s^x}(e^i - \frac{e^x + e^y}{2})$. If $d(x, L) < d(y, L)$, $f(v_L)$ is more dense closer to y, and if $d(x, L) > d(y, L)$, $f(v_L)$ is more dense closer to x.

Items 4 and 5 from Proposition 3 imply that the sure loser, L, cannot be extreme with respect to party competition. She can, however, be extreme with respect to the center of the distribution. In particular, L may be distant from the axis of party competition, but unless she is fully orthogonal and lies on the line $s^i = \frac{s^x + s^y}{2} + \frac{e^x - e^y}{s^y - s^x}(e^i - \frac{e^x + e^y}{2})$, she will still have preferences over x and y. So long as the density set of her voters, $f(v_L)$, is not symmetric over the dividing line between candidates x and y, she can asymmetrically influence the vote shares of candidates x and y. Moreover, those conditions also imply that the above line cannot be an element of M, the set of median lines.

2.5 Compromise Effects

Citizen candidate models typically offer a visual representation of candidates' positions, but voters translate the candidates' positions into their idiosyncratic issue space. Voters view candidates as bundles of issue positions in $\mathbb{R}^2$, and, for the purpose of framing effects analysis, have preferences exhibiting the compromise and attraction effects, as in Poterack (2015). Given a ballot of candidates, voters' preferences are influenced by the context created by the ballot. Specifically, the context is the vector of worst issue positions among the candidates running, called the *frame*. When comparing choices from ballots with the same frame, the decision makers' choices satisfy WARP; however, when comparing ballots with different frames, WARP may be violated. This implies the existence of a collection of complete and transitive preferences indexed by frames. The compromise and attraction effects are encompassed when lowering the frame changes the preference to one where the indifference curves are *rotated clockwise*, and moving the frame *left* gives new preferences such that the curves are rotated *counterclockwise*.

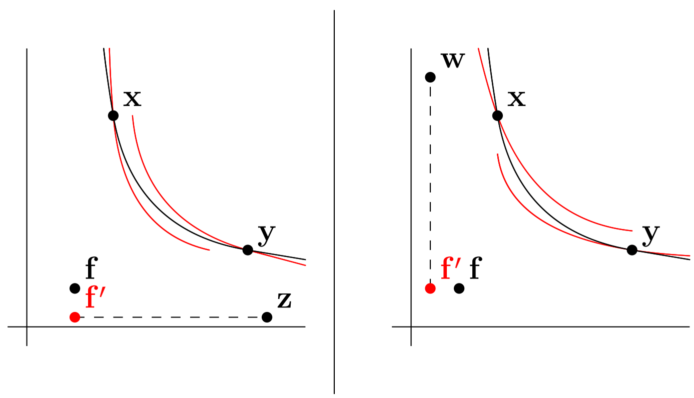

Figure 2·1: Indifference Curve Rotation in Response to Frame Changes

A voter's view of a candidate can be represented as a pair of negative numbers: the candidate's distance from the voter on economic issues, and the candidate's distance on social issues, both multiplied by negative one (because a larger distance makes for a less attractive candidate). In other words, voter a views candidate x as the bundle $(-|x_1 - a_1|, -|x_2 - a_2|)$.

This can be illustrated in the candidate space by drawing horizontal and vertical axes through the voter, and reflecting the candidates over these lines until they are all below and to the left of them. A ballot of candidates as follows would be translated to voter a's preference space as shown in Figure 2·2:

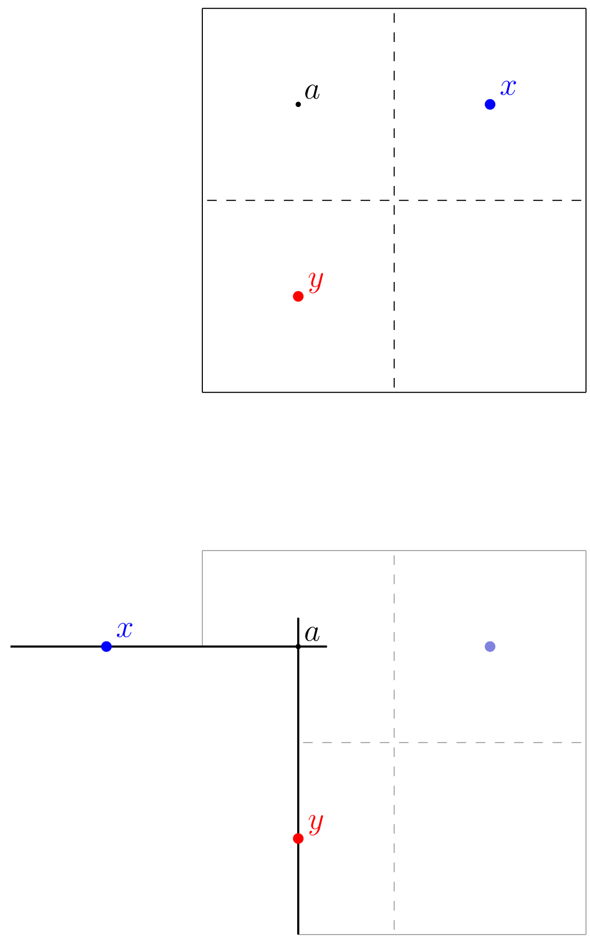

Figure 2·2: Translating from Candidate Space to Voter's Idiosyncratic Preference Space

To see the impact of the compromise and attraction effects, consider a voter space consisting of all possible combinations of economic and social positions. Suppose this space can be represented by $[0,1]^2$. Furthermore, suppose the underlying density is such that the yolk of the distribution is the point $(\frac{1}{2},\frac{1}{2})$. In other words, for both sets of issues, $\frac{1}{2}$ is the median position. Consider two candidates, x and y, which are both equidistant from the yolk and which both lie on a line passing through the yolk. Without loss of generality, suppose (as pictured above) x is in the upper right, and y is in the lower left.

Every voter in the upper right quadrant prefers x to y, because they are closer to x on both dimensions. Similarly, every voter in the lower left prefers y to x. If these quadrants contain the same measure of voters, this means x and y are tied when excluding the upper left and lower right quadrants. It is the votes in these quadrants which will break the tie.

Put another way, if the median is $(\frac{1}{2}, \frac{1}{2})$, $x = (\frac{1}{2} - n_1, \frac{1}{2} - n_2)$ $y = (\frac{1}{2} + n_1, \frac{1}{2} + n_2)$, and $a = (\frac{1}{2} - p_1, \frac{1}{2} + p_2)$, table 2.1 demonstrates how a views x and y:

	$n_1 > p_1$	$p_1 > n_1$
$n_2 > p_2$	$x = (-(n_1 + p_1), p_2 - n_2)$ $y = (p_1 - n_1, -(n_2 + p_2))$	$x = (-(n_1 + p_1), p_2 - n_2)$ $y = (n_1 - p_1, -(n_2 + p_2))$
$p_2 > n_2$	$x = (-(n_1 + p_1), n_2 - p_2)$ $y = (p_1 - n_1, -(n_2 + p_2))$	$x = (-(n_1 + p_1), n_2 - p_2)$ $y = (n_1 - p_1, -(n_2 + p_2))$

Table 2.1: Election view of voter a

This clearly indicates that a prefers x on one dimension, and y on the other. So it is possible a prefers x, or y. a may even be indifferent. This information is not useful.

However, consider a', a rotated 180° around the median ($a' = (m + p_1, m - p_2)$). a and a' make a line segment whose midpoint is also the midpoint of the line segment connecting x and y. This fact creates a relationship between a and a''s views of the candidates. a''s view of the candidates is in table 2.2:

a' has the exact opposite view of the candidates from a; a' views x in the same

	$n_1 > p_1$	$p_1 > n_1$
$n_2 > p_2$	$x = (p_1 - n_1, -(n_2 + p_2))$ $y = (-(n_1 + p_1), p_2 - n_2)$	$x = (n_1 - p_1, -(n_2 + p_2))$ $y = (-(n_1 + p_1), p_2 - n_2)$
$p_2 > n_2$	$x = (p_1 - n_1, -(n_2 + p_2))$ $y = (-(n_1 + p_1), n_2 - p_2)$	$x = (n_1 - p_1, -(n_2 + p_2))$ $y = (-(n_1 + p_1), n_2 - p_2)$

Table 2.2: Election view of voter a'

position that a views y, and vice versa. Furthermore, because they perceive a ballot with the same bundles, just swapped, they both have the same frame. In other words, if a prefers x, a' prefers y, and vice versa. Furthermore, if we assume the upper left and lower right quadrants have the same number of voters, then *there is an a' in the lower right for every a in the upper left.* Therefore, any votes x gets in the upper left are matched by votes for y in the lower right, and x and y tie.

Now, if a third candidate z enters, such that z is below and to the left of y, it is no longer the case that a and a' have equal and opposite views of the election. They both view z in a position where the other perceives no candidate. See Figure 2·3.

Because of this discrepancy, they perceive different frames. However, they each perceive a frame which asymmetrically benefits y, relative to the two candidate case. a perceives the frame as having shifted down, which benefits the lower candidate, while a' perceives the frame as shifting left, which benefits the leftmost candidate. This implies that while the introduction of z may induce a candidate in the upper left or lower right to switch from supporting x to supporting y, *none* of them will switch from supporting y to supporting x. z thus induces a gain in votes for y and reduction in votes for x. z also takes votes from y; some voters will prefer z to y. (If the indifference curves have a typical convex shape, z will only take votes from y, not x. z has incentive to enter if the votes they give y outweigh those taken.

This result applies whether or not z is on the same line as x and y. If a fourth candidate, w, enters on the line passing through z and the median, such that w and z are equidistant from the median, this makes the frames identical for a and a' again. Thus, they then again have equal and opposite preference over the candidates, and x and y again tie. This suggests a natural equilibrium has four candidates of this form.

To characterize a specific equilibrium, we impose the functional form for utility

$$U(e^x, s^x, \underline{e}^i, \underline{s}^i; e^i, s^i) = (|\underline{e}^i - e^i| - |e^x - e^i|)^{\frac{1}{2}} + (|\underline{s}^i - s^i| - |s^x - s^i|)^{\frac{1}{2}}$$

Moreover, assume that the support of F is given by $[0,1]^2$ and $F(.)$ is symmetric over $s^i = 1 - e^i$.

Proposition 4. There exists a four candidate equilibrium where candidates L_1, L_2, x, and y have ideal policies $(0,0)$, $(1,1)$, $(\frac{1}{2} - a, \frac{1}{2} - a)$, and $(\frac{1}{2} + a, \frac{1}{2} + a)$ respectively, and L_1 and L_2 lose with certainty, if and only if $\frac{3}{2}d(x,y) \geq c$, $b \geq 2c - d(x,y)$ and the following distributional conditions:

1. $v_x(L_1, L_2, x, y) = v_y(L_1, L_2, x, y)$.
2. $v_x(L_2, x, y) < v_y(L_2, x, y)$.
3. $v_y(L_1, x, y) < v_y(L_1, x, y)$.
4. $\nexists i$ such that $v_i(L_1, L_2, x, y, i) > \max_{j \in \{L_1, L_2, x, y\}} v_j(L_1, L_2, x, y, i)$.

The conditions on vote shares necessary to sustain equilibrium are characterized in Section B.1.2 of the appendix. Most noteworthy, in this equilibrium there are both upper and lower bounds on the number of voters who are in the quadrants of the policy space orthogonal to the axis of party competition. In essence, if voters are sufficiently correlated in their policy preferences to lie on the same dimension as party competition, they find it too easy to rank parties (one is strictly better than the other for a generic voter) and cannot be influenced by the compromise effect. If, on the other hand, there are sufficiently many voters who have preferences that are not represented well by candidates, there are too many voters who can be influenced by the compromise effect to sustain this equilibrium. In that case, a centrist entrant located at the yolk would obtain a plurality and win the election. For example, a uniform distribution over the space $[0,1]^2$ would not sustain this equilibrium, since an entrant at $(\frac{1}{2}, \frac{1}{2})$ would obtain a strict majority.

The qualitative properties of the equilibrium characterized in Proposition 4 are of particular interest. Previous models which generate sure losers either do so because they are unable to choose not to enter (Ansolabehere and Snyder 2000) or because they gain an advantage from cannibalizing the votes of their less preferred candidate (Solow 2015 and Section 2.4 of this paper). Instead, in this particular equilibrium, the sure losers are extreme with respect to party competition: they only receive votes from voters with whom share a second favorite candidate. Instead, their motivation to enter comes from changing the *perception* that moderate voters have of their preferred candidate.[5] Thus, candidates who are extreme with respect to party competition may still favorably affect their preferred moderate's chance of victory despite cannibalizing votes only from their preferred moderate. The second and third distributional conditions require that the votes accruing to candidate x from L_1 favorably altering how she is perceived are more numerous than the votes L_1 herself receives.

While the equilibrium in Proposition 4 features maximally extreme sure losers, there will generically exist equilibria with less extreme sure losers. For example, symmetrically perturbing L_1 and L_2 closer to the yolk, e.g. (ϵ, ϵ) and $(1-\epsilon, 1-\epsilon)$, may still be an equilibrium. The potential candidates at $(0,0)$ and $(1,1)$ may choose not to enter since they would frame L_1 and L_2 more attractively and potentially have a detrimental effect on candidates x and y. Nevertheless, if there exists a linear

[5] There is suggestive experimental evidence that this is the case. Eric Loepp, writing in the Washington Post, cites preliminary results of an experiment he is conducting that suggests that Republican voters change their evaluation of a moderate candidate's ideology when paired with an extreme conservative candidate. https://www.washingtonpost.com/news/monkey-cage/wp/2016/03/16/trump-changes-how-voters-view-the-other-republican-candidates-heres-how/, accessed March 30, 2016.

equilibrium with extremist sure losers who are not maximally differentiated, then there surely exists a linear equilibrium with maximally differentiated sure losers. The intuition is straightforward: all potential candidates who are extreme with respect to party competition on the line have utility of candidates x and y winning the election the same as all other potential candidates in their neighborhood. Thus, if an extreme entrant who is not maximally differentiated finds it optimal to enter as a sure loser, the payoff conditions will also be satisfied for a maximally differentiated extremist. The maximally differentiated extremist, however, provides a more favorable frame to their preferred moderate (since they are further away), and steals strictly fewer votes. Therefore, the distributional conditions summarized in Section B.1.2 are necessary conditions for any equilibria in this class.

2.6 Conclusion

In this paper, we develop a multidimensional citizen-candidate model of entry into plurality elections. We show that limiting the degree of commitment available to potential candidates restores pure strategy Nash equilibria which do not exist in a Hotelling-Downs model of multidimensional spatial competition. Moreover, expanding the issue space past one dimension provides novel incentives for candidates. We characterize the first multicandidate equilibria with a sure winner in a political entry game with single-peaked preferences. This equilibrium has normatively undesirable properties which occur even in examples without single-peaked preferences. In particular, the certain victor is a Condorcet loser among the set of candidates for office.

We then utilize the multiple dimensions to study two well-documented violations of the Weak Axiom of Revealed Preference: the attraction and compromise effects. We model these effects via frames of reference and show that the behavior of extremists can have asymmetric effects which favor their preferred moderate candidate.

Moreover, we show that with the right distribution of preferences there may exist an equilibrium where extremists enter despite losing with certainty solely to enhance the probability of victory of their preferred moderate through framing effects.

We do not, however, provide a full characterization of all equilibria of this game, especially in the setting with behavioral voters. Of particular interest in further research would be modifying the frame from the composite minima we study to an alternative formulation that does not require boundedness of the issue space. While equilibria with extreme sure losers who are not on the boundary do exist in our model, expanding the issue space will always give greater incentive for sure losers on the boundary. This is an undesirable feature of our formulation.

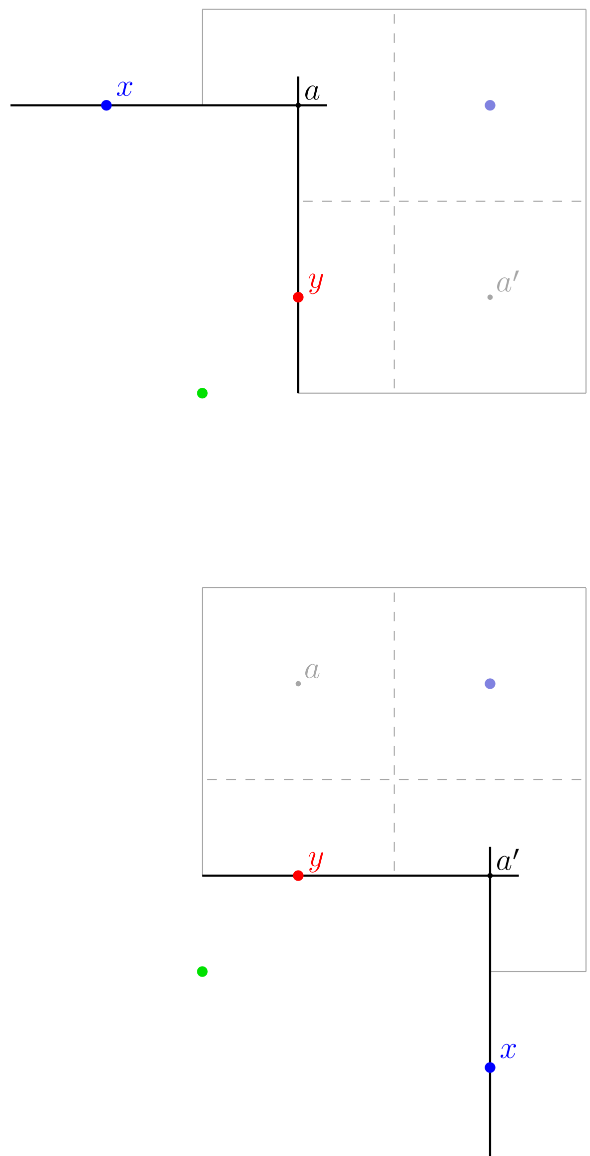

Figure 2·3: Asymmetric Framing Effects

Chapter 3

Giving the Gift of Guilt Avoidance

3.1 Introduction

One remarked upon result from economic theory is that a gift in kind cannot increase surplus more than a gift in cash. The popularity of giving a gift card, allowing the recipient to purchase whatever they desire from a given merchant, partially supports this theory, as a gift card is closer to cash than a gift of an actual good; of course, the same theory again suggests that cash is better still. Is it possible for a gift card to increase its recipient's surplus by *more* than a gift of cash in the same amount? This is both a question about the deadweight loss of gift giving, and about preferences over menus, specifically the preference for commitment to a smaller menu. A gift of cash is equivalent to a menu of all goods that can be purchased with the cash, and the gift card is a smaller menu, restricted only to goods at the merchant issuing the card.

There is extensive work in the literature on preferences for smaller menus, most notably in the work on temptation beginning with Gul and Pesendorfer (2001). They take a set of lotteries X, and define a preference $\succsim$ over the set of *menus* $\mathcal{M}$, i.e., the collection of all non-empty compact subsets of X. Their decision makers face a struggle between normative and temptation preferences: what they think they *should* choose, and what they might be *tempted* to choose. For example, a dieter may have a normative preference to eat salad, but be tempted to eat a burger (and potentially give in to this temptation). Given a normative preference for menu A over B, the decision maker weakly prefers to *commit* to menu A, rather than face it alongside B, which may contain tempting elements. Even if the decision maker does not succumb to temptation, they still may pay a self-control cost for resisting it. Gul and Pesendorfer (2001) capture this with their main axiom, *Set-Betweenness*: given menus $A, B \in \mathcal{M}$, if $A \succsim B$, then $A \succsim A \cup B \succsim B$.

Temptation therefore provides a rationale for why a decision maker would prefer to commit to a menu of normatively preferred goods. However, could a decision maker prefer to commit to a normatively *dispreferred* menu, rather than face a larger menu? One possible reason for this would be the presence of guilt. Consider the point of view of a dieter. He would prefer to eat a salad $\{s\}$, but if he's at a restaurant with a burger $\{b\}$ on the menu, he will be tempted to order the burger instead. *Set-Betweenness* implies the preference $\{s\} \succsim \{s, b\} \succ \{b\}$; i.e., he would prefer a menu with just salad to a menu with both salad and burger, so that he is not tempted by the burger. Importantly, he would weakly prefer the menu with both salad and a burger to the menu with just a burger, as it provides the possibility of ordering salad, and even if temptation is succumbed to, ordering a burger off the menu with both is *no worse* than ordering off the menu with burger alone.

This may not hold in a world with guilt. In a world with guilt, it is possible that in addition to the self control cost which makes it strictly worse to order a salad off the menu $\{s, b\}$ than off the menu $\{s\}$, there is also a *guilt cost* which makes it strictly worse to order a burger off the menu $\{s, b\}$ than to order a burger off the menu $\{b\}$. In this instance, we may observe the preferences $\{s\} \succsim \{b\} \succsim \{s, b\}$; the menu $\{s, b\}$ is the least preferred because it leaves the decision maker in a double bind, doomed to pay either a self control cost or a guilt cost, no matter which he chooses. Guilt behavior therefore can be expressed by the axiom *Partial Set-Betweenness*, from Kopylov (2007). The axiom weakens *Set-Betweenness* by allowing for the preference

$A \gtrsim B \succ A \cup B$. If menu A is preferred to menu B, the decision maker always weakly prefers to commit to A rather than choose from the union of both menus. However, unlike in *Set Betweenness*, the decision maker does not necessarily prefer the union of both menus to comitting to B, *unless* A and B both contain the same normatively best element. Guilt arises from rejecting a normatively preferred element in favor of a tempting one. Ordering a burger off of a menu that also includes salad incurs no more guilt than ordering one off a menu that also includes salad and a hot dog; in both cases, the guilt arises from spurning the salad to order a less healthy item. This axiom applies well to the observation at hand. If x represents a menu of normatively preferred necessities, and y represents a gift card for tempting indulgences, $\{x, y\}$ may represent cash which can be spent on either, and the desired menu preference is $\{x\} \gtrsim \{y\} \gtrsim \{x, y\}$. Committing to y removes the guilt cost of not choosing to buy necessities. As such, I refer to the act of committing to a normatively dispreferred menu as "guilt avoidance."

However, this suggests decision makers might immediately use their cash to buy gift cards, thereby allowing guilt free consumption. This seems self-defeating, as the very act of trying to *avoid guilt* by purchasing the gift card should itself *incur guilt*. Therefore, the desired model would indicate a preference for *receiving* a gift card, but not for choosing one for one's self.

This fits well within the realm of gift giving. Clearly, when the benefit from *receiving* a good differs from the benefit of *choosing* the good for one's self, there is the potential for benefit from gift giving. This arises naturally in the context of guilt. A decision maker cannot avoid guilt by choosing a limited menu for herself, as this very act of guilt avoidance itself incurs guilt. However, if she receives the menu unprompted from another, she has received the gift of guilt avoidance.

To capture this, I propose a multiperiod model, where in the first period the decision maker is *gifted* a menu of menus, and in the second period *chooses* from this menu. Second period preferences are revealed by choice; first period preferences are revealed by amounts of cash the decision maker would be indifferent to. In other words, if the decision maker strictly prefers receiving the menu of menus A to receiving menus of menus B, this means that when in possession of A, the amount of cash they would need to receive to trade A is greater than the amount they'd need to trade B.

Why use such an arcane procedure for revealed preference? The very behavior of interest seems to suggest an agent who, when in possession of $\$n$, would not spend it on a gift card, but when in possession of an $\$n$ gift card, would want *more* than $\$n$ to trade it. This seems to be a distinction between choice behavior and trade behavior. I argue this stems from the observation that people feel more moral culpability from *action* than from *inaction*, even when the distinction between the two is slight. For a decision maker to purchase herself a gift card is an action, which would inspire guilt, and defeat the purpose of buying the gift card. However, to refuse a trade for an already possessed card is inaction, which does not inspire moral culpability, and results in no guilt.

Formally, the analysis proceeds with a three period model, inspired by Noor and Ren (2015). There is a set of lotteries X, a set of menus of lotteries $\mathcal{M}_1$, i.e., the set of all non-empty compact subsets of X, and a set of menus of menus $\mathcal{M}_0$, the set of all non-empty compact subsets of $\mathcal{M}_1$. The primitive is a preference $\succsim$; this primitive can be interpreted as arising from a dollar amount associated with each menu of menus, which represents the minimum amount the decision maker is willing to accept in trade for the menu of menus, given that they have received it as a gift. Given $A, B \in \mathcal{M}_0$, $A \succsim B$ if the dollar amount associated with A is greater than or equal to that associated with B. Preferences over these menus of menus reflect the preference for commitment to normatively dispreferred items in order to avoid guilt. In the second period, the decision maker *chooses* a menu from the menu of menus, and because they are *choosing*, they do not have preference for commitment to normatively dispreferred items, because this commitment in itself would incur guilt. Singleton menus of menus are a commitment to the menu for the last period, which effectively removes the interim period. In the interim period, menus are chosen from the menu of menus according to the preferences over commitment to menus. Because menus are chosen in this period, and this choice is active, these preferences do not reflect the desire to commit to a normatively dispreferred menu for purposes of guilt avoidance. The desired behavior, therefore, is that while preferences over all menus of menus obey *Partial Set-Betweenness*, preferences over singleton menus of menus obey *Set-Betweenness.* I propose coupling *Partial Set-Betweenness* with the axiom *Singleton Set-Betweenness* which captures this behavior.

If a preference $\succsim$ over $\mathcal{M}_0$ obeys *Weak Order, Continuity, Independence* (as all the preceding axioms are typically defined), *Partial Set-Betweenness*, and *Singleton Set-Betweenness*, then there exists a representation of the following form:

$$\begin{aligned} W(A) &= \max_{a\in A}\left[U(a) - \max_{b\in A}\left(V(b) - V(a)\right) - \kappa \max_{c\in A}\left(U(c) - U(a)\right)\right] \\ \text{where } U(a) &= \max_{x\in a}\left[u(x) - \max_{y\in a}\left(v(y) - v(x)\right)\right] \end{aligned}$$

where u, v, U, and V are positive, linear functions. Clearly, this functional form is related to GP, but with the additional term $\kappa \max_{c\in A}(U(c) - U(a))$, representing guilt. This term is found in Kopylov's working paper, and represents the guilt cost from not choosing the normatively best element in a menu. It is not found in the $U(a)$ term, because this represents the decision maker's preferences over active choice, where guilt avoidance is not a factor.

3.2 Literature Review

This paper borrows from Kopylov (2007), notably in the concept of a preference for smaller menus as a means of guilt avoidance. That paper, however, does not address the notion that guilt avoidance causes guilt; it is a single period model, with no distinction between action and inaction.

Noor and Ren (2015) is based around active choice, and embraces the notion that guilt avoidance causes guilt; because of this, they reject that a decision maker can prefer a smaller menu for reasons of guilt avoidance. As such, they maintain *Set-Betweenness*, and do not allow for the possibility of preferring a menu restricted to normatively dispreferred goods. However, my point is that when considering inactive choice, guilt avoidance does not necessarily lead to guilt, because the decision maker lacks moral culpability. This can be reflected by an alternative interpretation of revealed preference based on trades for menus of menus one is already in possession of.

This paper also has implications for work on the deadweight loss of gift-giving; the conclusions reached here allow for the possibility of surplus from gift-giving. There have been many papers using survey and experimental data to measure the deadweight loss (or surplus) of gift giving; these include a series beginning with Waldfogel (1993), which saw responses from Solnick and Hemenway (1996) and List and Shogren (1998). They have reached differing conclusions, some showing a deadweight loss from gift giving, and others showing a surplus. These papers propose asymmetric information as a potential source for surplus: a gift giver who knows more about a recipient's preferences than the recipient can well increase the recipient's surplus through a gift in kind. There is no suggestion of guilt avoidance as a potential source of surplus. There is also work on theoretical frameworks allowing for surplus, including Solow (1993) and Thaler (1985). Neither paper explicitly models guilt. Solow suggests the possibility of externalities which accrue to the gift-giver. Thaler has a model of mental accounting, wherein decision makers artificially constrict themselves from allocating additional cash gifts to budgets for goods they may desire to receive as gifts. This idea of mental accounting could be motivated by guilt, though Thaler does not explicit state this. The representation in his paper is distinct from this paper, and he does not provide a behavioral foundation.

3.3 The Model

3.3.1 Primitives

Consider a set of goods X, a convex subset of a linear space. Mixtures $\alpha x + (1-\alpha)y$ are well defined for all $\alpha \in [0,1]$ and $x, y \in X$. Use metric d such that

1. X is compact

2. mixture operation mapping $[0,1] \times X \times X \to X$ is continuous; i.e., $\alpha x + (1-\alpha)y$ is continuous in α for all x, y

3. For each $\alpha \in [0,1]$ and $x, y, x', y' \in X$

$$d(\alpha x + (1-\alpha)x', \alpha y + (1-\alpha)y') \leq \max\{d(x,y), d(x',y')\}$$

This general specification applies to many settings used in other decision theory papers. For example, these conditions are satisfied when X is the convex compact subset of a normed linear space with the metric induced by the norm, as used by Dekel et al. (2001), or the class of all Borel probability measures on a compact metric Z with the Prohorov metric d of the weak convergence topology, as used by Gul and Pesendorfer (2001).

Let $\mathcal{M}_1$ be the set of all non-empty compact subsets of X, i.e., the set of menus.

For each $a, b \in \mathcal{M}_1$, define mixtures and the Hausdorff metric μ_1 in $\mathcal{M}_1$ by

$$\begin{aligned}
\alpha a + (1-\alpha)b &= \{\alpha x + (1-\alpha)y | x \in a,\ y \in b\} \\
\mu_1(a,b) &= \inf\{\varepsilon > 0 | b \subset a^\varepsilon \text{ and } a \subset b^\varepsilon\}
\end{aligned}$$

where a^ε and b^ε are the ε-neighborhoods of the sets a and b in the metric space (X, d). According to Kopylov (2009b), $\mathcal{M}_1$ is a compact metric space with continuous mixture operation.

Now consider $\mathcal{M}_0$, the set of all non-empty compact subsets of $\mathcal{M}_1$. $\mathcal{M}_0$ is the set of "menus of menus". Endow $\mathcal{M}_0$ with a mixture operation and metric defined analogously as those on $\mathcal{M}_1$.

There is a three period decision making process. In period 0, the decision maker is gifted a menu of menus $A \in \mathcal{M}_0$. In period 1, the decision maker chooses a menu $a \in A$. In period 2, the decision maker chooses and consumes a good $x \in a$.

The guilt behavior studied in this paper reflects decision makers who would prefer to receive a gift that they would not choose to buy themselves. The primitive for this model is a preference $\succsim$. I interpret this preference as being revealed by a series of dollar amounts associated with each menu of menus. Each dollar amount represents the minimum the decision maker would accept in trade for a menu of menus, given that they are already in possession of it. The difference between receiving a gift and purchasing a menu is moral culpability. The former is passive and the latter is active. The decision maker feels guilt over action, but not over inaction. If the decision maker purchased herself a gift card, this attempt at guilt avoidance would itself inspire guilt which cancels out any guilt avoidance benefit. However, to receive such a gift does not inspire guilt, allowing her to indulge in guilt avoidance. This is what motivates the primitive. By asking to trade a menu of menus already in possession, the desire for guilt avoidance can be isolated. In rejecting cash in trade, the decision maker feels morally justified and guilt free. As such, her preference should be interpreted as normative. $A \gtrsim B$ is interpreted as meaning that the dollar amount associated with A is greater than or equal to the dollar amount associated with B. Given a utility function representing $\gtrsim$, it can be interpreted as follows: the dollar amounts associated with each menu of menus are an increasing function of the utility value of each menu of menus. In period 1, the decision maker chooses the highest ranking menu in their menu of menus according to the preference $\gtrsim$ over singleton menus of menus. Because the decision maker is making an active choice, they cannot successfully avoid guilt, and these preferences do not represent the desire for commitment to normatively dispreferred menus.

3.3.2 Representation

The desired representation is of the following form:

$$\begin{aligned} W(A) &= \max_{a \in A}\left[U(a) - \max_{b \in A}(V(b) - V(a)) - \kappa \max_{c \in A}(U(c) - U(a))\right] \\ \text{where } U(a) &= \max_{x \in a}\left[u(x) - \max_{y \in a}(v(y) - v(x))\right] \end{aligned}$$

where u, v, and V are positive, linear functions[1]. This is a Gul and Pesendorfer (2001) representation applied to multiple periods, with the addition of the term $-\kappa \max_{c \in A}(U(c) - U(a))$, representing the desire for guilt avoidance. This term gives the decision maker a "guilt cost" from not choosing the most normatively preferred menu from a menu of menus, similar to the "self control cost" term $-\max_{b \in A}(V(b) - V(a))$ which imposes a cost of not choosing the most tempting menu. As the "self control cost" creates a desire to commit to a smaller menus of menus, restricted to normatively preferred menus, the "guilt cost" creates a desire to commit to a smaller menu

[1]Note that V is independent of v.

of menus which *excludes* more normatively preferred menus.

Given a comparison between choosing menu a from two menus of menus $A, B \in \mathcal{M}_0$ such that they both contain the same normatively best menu, the "guilt cost" term will be identical for both, meaning the comparison will reduce to a Gul and Pesendorfer (2001) representation. If the choice a is not contained in both menus, the term is not meaningful. This will have implications for the behavioral axioms used to explore guilt avoidance, discussed in the next subsection.

Note that the function $U(a)$ represents the normative ranking of menus, even though it contains self control costs, just as in Gul and Pesendorfer (2001). Furthermore, the function $W(A)$ represents the normative ranking of menus of menus, even though it contains both self control and guilt costs. As mentioned in the previous section, the preference $\gtrsim$ over menus of menus reflects a desire for guilt avoidance, but is still normative, as the decision maker feels morally justified in her guilt avoidance.

Numerical Example

The fact that this representation captures the desired behavior can be shown in the following example. Suppose a decision maker desires necessities, x, and luxuries, y. The menu of menus $\{\{y\}\}$ represents the gift of a gift card which can only be spent on luxuries, and the menu of menus $\{\{x, y\}, \{y\}\}$ represents the gift of cash, which, in the interim, can be kept as cash, or spent on a gift card. Suppose there is no gift card to commit to spending on necessities. If the decision maker keeps the cash as cash, she is choosing the menu $\{x, y\}$ from the menu of menus they have been gifted. If she spends it on a gift card, she has chosen the menu $\{y\}$.

Suppose $u(x) = 3$, $v(y) = 2$, $u(y) = v(x) = 0$, $\kappa = 2$. In other words, x is normatively preferred, y is tempting, and the decision maker experiences guilt. Suppose $V(\{y\}) = 2$ and[2] $V(\{x, y\}) = 0$. This gives rise to the following utility values:

[2]Noor and Ren (2015) give axioms which guarantee for a similar model that $V(a) =$

$$W(\{\{y\}\}) = U(\{y\}) = u(y) = 0 \tag{3.3.1}$$

$$W(\{\{x,y\}\}) = U(\{x,y\}) = u(x) - (v(y) - v(x)) = 1 \tag{3.3.2}$$

$$W(\{\{x,y\},\{y\}\}) = U(\{x,y\}) - (V(\{y\}) - V(\{x,y\})) - \kappa(0) = -1 \tag{3.3.3}$$

The gift of the gift card returns a utility of 0, which makes it preferred to the gift of cash $\{\{x,y\},\{y\}\}$, which returns a utility of -1. However, if the decision maker is given cash, then in the interim, she faces a choice between spending the cash gift on a gift card, i.e., choosing $\{y\}$, or keeping it as cash, i.e., choosing $\{x,y\}$. This decision is made according to the preference over the relevant singleton menus of menus, which shows that keeping the cash as cash is preferred, returning a utility of 1. Receiving cash is worse than keeping cash, because receiving cash includes the option to purchase a gift card, leaving the decision maker in the double bind of being *required* to face either a guilt cost or a self control cost.

3.3.3 Axioms

Axiom 1 (*Weak Order*)**.** $\succsim$ is complete and transitive.

Axiom 2 (*Continuity*)**.** For all menus of menus $A \in \mathcal{M}_0$, the sets $\{B \in \mathcal{M}_0 |\ B \succsim A\}$ and $\{B \in \mathcal{M}_0 | A \succsim B\}$ are closed.

Axiom 3 (*Independence*)**.** For all $\alpha \in [0,1]$ and menus of menus $A, B, C \in \mathcal{M}_0$,

$$A \succ B \Rightarrow \alpha A + (1-\alpha)C \succ \alpha B + (1-\alpha)C$$

Weak Order, *Continuity*, and *Independence* are standard conditions; the justification for the use of *Independence* found in Gul and Pesendorfer (2001) and Dekel et al. (2001) applies here as well.

$\max_{x \in a}[v(x) - \max_{y \in a}(u(y) - u(x))]$, which is consistent with the values given here.

Gul and Pesendorfer (2001) use the axiom *Set-Betweenness*, which states that for any menus $a, b \in \mathcal{M}_1$, $a \succsim b \Rightarrow A \succsim A \cup B \succsim B$. In the present model, the presence of guilt is evidenced by the desire to commit to a normatively dispreferred menu of menus, rather than face a larger menu of menus containing it and a normatively preferred menu of menus. Clearly, *Set-Betweenness* rules out this behavior, and needs to be replaced with a weakened axiom.

Because commitment to a normatively dispreferred menu is motivated by guilt avoidance, it should only be desired when the smaller, normatively dispreferred menu of menus excludes the normatively best menu from the larger menu of menus. For any $c \in \mathcal{M}_1$, define $\mathcal{M}_c = \{A \in \mathcal{M}_0 | c \in A \text{ and } \{c\} \succsim \{b\} \text{ for all } b \in A\}$. This is the set of all menus of menus with the same normatively best menu c. As there is no motivation for guilt avoidance by choosing a smaller menu of menus from $\mathcal{M}_c$ over a larger menu of menus also in $\mathcal{M}_c$, preferences over these menus of menus should not reflect guilt avoidance, and not violate *Set-Betweenness*. This behavior is captured by the following axiom, due to Kopylov (2007):

Axiom 4 (*Partial Set-Betweenness*)**.** For all $c \in \mathcal{M}_1$ and menus of menus $A, B \in \mathcal{M}_0$,

1. *if* $A \succsim B$, *then* $A \succsim A \cup B$
2. *if* $A \succsim B$ *and* $A, B \in \mathcal{M}_c$, *then* $A \cup B \succsim B$.

As mentioned above, choices made by the decision maker in period 1 are represented by the preference $\succsim$ over singleton menus. The decision maker in period 1 is engaged in active choice, and thus susceptible to feeling guilt for engaging in guilt avoidance. As such, they do not exhibit a preference for guilt avoidance in this period; the guilt they would feel over actively engaging in guilt avoidance renders such action meaningless. This behavior is captured by the following axiom:

Axiom 5 (*Singleton Set-Betweenness*)**.** For all singleton menus of menus $\{a\}, \{b\} \in \mathcal{M}_0$, $\{a\} \succsim \{b\} \Rightarrow \{a\} \succsim \{a \cup b\} \succsim \{b\}$.

The interpretation of the axioms is that temptation is experienced in both the interim and final stages, and guilt is experienced in the interim stage. Given the preference $\{\{x\}\} \succ \{\{y\}\}$ over goods $x, y \in X$, the axioms allow for the preference $\{\{x\}\} \succ \{\{x, y\}\}$, which indicates a desire for commitment to avoid temptation in the final stage, as this is when the choice from the menu $\{x, y\}$ would be made. Similarly, given the preference $\{a\} \succ \{b\}$ over menus $a, b \in \mathcal{M}_1$, the axioms allow for the preference $\{a\} \succ \{a, b\}$. This indicates a desire for commitment to avoid temptation in the *interim* stage, as this is when the choice from the menu of menus $\{a, b\}$ would be made. Preferences consistent with the axioms are also naturally interpreted as showing guilt to be experienced in the interim stage, because again given the preference $\{a\} \succ \{b\}$ over menus $a, b \in \mathcal{M}_1$, the axioms allow for the preference $\{b\} \succ \{a, b\}$. This must indicate a desire to avoid guilt in the *interim* stage. When facing the menu of menus $\{a, b\}$ in the interim stage, the decision maker will choose a (as $\{a\} \succ \{b\}$), and facing a in the final stage does not inspire guilt in that stage.

The model does not make an explicit claim about whether guilt is experienced in the final stage. Guilt is reflected in behavior by guilt avoidance, the desire for commitment to normatively dispreferred menus. However, guilt in the final stage cannot inspire this behavior in the interim stage, because this behavior would in and of itself cause guilt. The decision maker is sophisticated enough to recognize that if she makes any attempt to avoid guilt by choosing a restricted menu in the interim period, she has committed an act for which she is morally culpable, and this cancels out any gains from guilt avoidance. As such, there is no behavioral expression of guilt experienced in the final stage. While certain preferences may only indicate interim period guilt, as demonstrated in the previous paragraph, other preferences can be interpreted as revealing guilt in either stage. For example, given goods $x, y \in X$ such that $\{\{x\}\} \succ \{\{x, y\}\} \succ \{\{y\}\}$, the axioms allow for the preference

$\{\{y\}\} \succ \{\{x,y\},\{y\}\}$. This preference can be interpreted as a desire to avoid guilt in the interim *or* in the final stage. The decision maker may pay a psychic guilt cost from facing the menu $\{x,y\}$ in the final stage as opposed to menu $\{y\}$, but this can only be behaviorally expressed in her preferences over menus of menus.

The preceding axioms allow for the following theorem:

Theorem 1 (Multi-Period Guilt Representation Theorem)**.** A preference $\succsim$ over the set of menus $\mathcal{M}_0$ satisfies *Order, Continuity, Independence, Partial Set-Betweenness,* and *Singleton Set-Betweenness* if and only if it has a representation of the following form:

$$W(A) = \max_{a \in A} U(b) - \max_{b \in A} [V(b) - V(a)] - \kappa \max_{c \in A} [U(c) - U(a)]$$

where u, v, U and V are linear, positive functions.

Proof. See appendix C.1. The proof is an extension of the proof from Kopylov (2007) to multiple periods. □

3.4 Conclusion

Guilt is a feature of human psychology which has many behavioral implications; one of them is a desire to limit one's options in order to avoid guilt. Decision makers may gain a benefit from being forced into choosing a tempting option which their normative preferences steer them away from, and this benefit can be captured in the representation offered by this paper. Previous attempts to model this have run into the difficulty that guilt avoidance causes guilt, and need to find various ways to deal with this. This paper proposes a novel definition of revealed preference in an attempt to avoid this difficulty. This definition of revealed preference may have application to other observations from decision theory, such as the endowment effect.

This benefit from commitment also has implications for gift giving. In order for the possibility of gains from gift-giving to exist, there must be something the giver can offer the recipient that the recipient cannot given themselves. Other papers have focused on the possibility of the giver having additional information; this paper focuses on the gift of guilt avoidance. Guilt afflicted decision makers are explicitly unable to avoid their own guilt; they need another to avoid it for them. This allows for gains from gift giving, as the gift giver is able to give a gift the decision maker cannot give themselves.

While this paper focuses on a decision making procedure over a small number of periods, an infinite horizon model could allow for an alternative conceptualization of the behavior of interest. A future oriented decision maker may underinvest in current consumption, and have an interest in commitment to such consumption, even if it is normatively dispreferred. This is a promising area for future research, particularly as it may be compatible with a more typical revealed preference approach. I chose a different approach because I felt it more clearly isolated the particular behavior of interest, especially in the context of a gift-giving situation, which does not have an obvious infinite horizon extension.

Appendix A

Appendices to The Compromise and Attraction Effects Through Frame Preferences

A.1 The n-attribute Case

For simplicity, the main body of the paper assumes there are only two attributes along which goods are judged. However, the model extends naturally to the n-attribute case, with minimal modification. Suppose the set of goods is $\mathbb{R}^n$. As before, define attribute preferences $\succsim_1, \ldots, \succsim_n$. The natural extension of the frame definition is $\mathbf{f}(S) = (\min_{\mathbf{x} \in S} x_1, \ldots, \min_{\mathbf{x} \in S} x_n)$. Having defined the frame, it is now easy to define the frame preference by

$$\mathbf{x} \succsim^{\mathbf{f}} \mathbf{y} \Leftrightarrow \exists\ S \in \mathcal{S} \text{ such that } \mathbf{x}, \mathbf{y} \in S,\ \mathbf{f}(S) = \mathbf{f},\ \mathbf{x} \in C(S)$$

As before, $\succsim^{\mathbf{f}}$ is only defined on

$$A^{\mathbf{f}} \equiv \{\mathbf{x} \in X | x_i \geq f_i\ \forall\ i \in \{1, \ldots, n\}\} \tag{A.1.1}$$

Finally, the definition of the pairwise preference $\succsim^*$ is unchanged: $\mathbf{x} \succsim^* \mathbf{y} \Leftrightarrow \mathbf{x} \in C(\{\mathbf{x}, \mathbf{y}\})$.

The properties have trivial extensions to the n-attribute case, with the exception of Property 2, which poses more of a challenge. Recall that in the $\mathbb{R}^2$ case, Property 2 is a statement about changes in the slope of the indifference curve; therefore, in the

$\mathbb{R}^n$ case, it is a statement about changes in the norm of the hyperplane tangent to the indifference surfaces.

By Property 2, lowering the frame makes the indifference curves steeper. Another way to express that is when the frame is decreased in the second component, a vector perpendicular to an indifference curve at a given point will also decrease in the second component. Therefore, the n-dimensionally equivalent statement is that when the frame is decreased in the ith component, the norm of the hyperplane tangent to the indifference surface *also* decreases in the ith component.

Define $u^{\mathbf{f}}(\mathbf{x}) \equiv U(\mathbf{x}, \mathbf{f})$. $\nabla u^{\mathbf{f}}(\mathbf{x}) = \langle u_1^{\mathbf{f}}(\mathbf{x}), \ldots, u_n^{\mathbf{f}}(\mathbf{x})\rangle$ is the norm of the hyperplane tangent to the indifference surface on which $\mathbf{x}$ lies. The desired property is that if $\mathbf{f}' \equiv (\mathbf{f}_1, \ldots, \mathbf{f}_i - \delta, \ldots, \mathbf{f}_n)$, for some $\delta > 0$, then $\exists\ \varepsilon > 0$ such that

$$\frac{\nabla u^{\mathbf{f}'}(\mathbf{x})}{\|\nabla u^{\mathbf{f}'}(\mathbf{x})\|} = \frac{\langle u_1^{\mathbf{f}}(\mathbf{x}), \ldots, u_i^{\mathbf{f}}(\mathbf{x}) - \varepsilon, \ldots, u_n^{\mathbf{f}}(\mathbf{x})\rangle}{\|\langle u_1^{\mathbf{f}}(\mathbf{x}), \ldots, u_i^{\mathbf{f}}(\mathbf{x}) - \varepsilon, \ldots, u_n^{\mathbf{f}}(\mathbf{x})\rangle\|}$$

i.e, the norm of the new hyperplane tangent to the indifference surface on which $\mathbf{x}$ lies after the frame is lowered on the ith component to $\mathbf{f}'$ overlaps the old one decreased on the ith component by some amount.

As with the properties, the axioms remain either entirely unchanged, or have trivial extensions to the n-attribute case, with the exception of *Compromise/Attraction Monotonicity*. The extension of this axiom requires some care; however, upon reflection it should seem quite natural and intuitive, and clearly equivalent to the n-dimensional statement of Property 2.

- *Compromise/Attraction Monotonicity*: Given $\mathbf{x} \sim^{\mathbf{f}} \mathbf{y}$, define $A, B \subset \{1, \ldots, n\}$ s.t. $i \in A \Rightarrow \mathbf{x} \succ_i \mathbf{y}$, $j \in B \Rightarrow \mathbf{y} \succ_j \mathbf{x}$.

  1. For each $i \in A$,

     – $\mathbf{y} \succ^{(f_{-i}, f_i')} \mathbf{x}\ \forall\ f_i' < f_i$

     – $\mathbf{x} \succ^{(f_{-i}, f_i'')} \mathbf{y}\ \forall\ f_i'' \in (f_i, y_i]$

2. For each $j \in B$,
   - $\mathbf{x} \succ^{(f_{-j}, f'_j)} \mathbf{y} \ \forall \ f'_j < f_j$
   - $\mathbf{y} \succ^{(f_{-j}, f''_j)} \mathbf{x} \ \forall \ f''_j \in (f_j, x_j]$
3. For each $k \in (A \cup B)^c$,
   - $\mathbf{x} \sim^{(f_{-k}, f'_k)} \mathbf{y} \ \forall \ f'_k$

Now with the axioms sorted, proceed as before, proposing a representation. Define $v_i(\mathbf{x}, \mathbf{f})$ as the solution to

$$\mathbf{x} \sim^{\mathbf{f}} (f_1, \ldots, f_{i-1}, v_i(x_1, \ldots, x_n, f_1, \ldots, f_n), f_{i+1}, \ldots, f_n)$$

v_i is a representation. It is strictly increasing and continuous in x_j, $\forall \ j$. v_i is decreasing and continuous in f_i. Given $f'_i > f_i$, define $\mathbf{f'} \equiv (f_1, \ldots, f'_i, \ldots, f_n)$, and

$$(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n) \succ^{\mathbf{f'}} (f_1, \ldots, v_i(\mathbf{x}, \mathbf{f'}), \ldots, f_n)$$

This is because it must be the case that $v_i(\mathbf{x}, \mathbf{f}) \succ_i \mathbf{x}$. Now, consider $f_i^m \to f_i$. Without loss of generality, suppose this is increasing. Define $\mathbf{f}^m \equiv (f_1, \ldots, f_i^m, \ldots, f_n)$. $\mathbf{f}^m \to \mathbf{f}$. Because v_i is decreasing in f_i,

$$v_i(\mathbf{x}, \mathbf{f}) \leq v_i(\mathbf{x}, \mathbf{f}^m) \ \forall \ m$$

Because $v_i(\mathbf{x}, \mathbf{f}^m)$ is bounded from below, $\exists \ \inf_m \{v_i(\mathbf{x}, \mathbf{f}^m)\}$, and $v_i(\mathbf{x}, \mathbf{f}) \leq \inf_m \{v_i(\mathbf{x}, \mathbf{f}^m)\}$. If this inequality were strict, it would violate *Frame Continuity*, so it must hold with equality, which implies continuity in f_i.

Now consider $\{(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n)\}_{i=1}^n$. All of these points are $\mathbf{f}$-indifferent to $\mathbf{x}$, so they are all $\mathbf{f}$-indifferent to each other. Furthermore, for any two of them, the frame of just the pair is $\mathbf{f}$. So any two of the points are pairwise indifferent. By *Pairwise Weak Order*, $\exists \ u^*$ representing the pairwise preference. u^* can show v_i is decreasing and continuous in f_j.

Consider $\{f_j^p\}_{p=1}^{\infty} \to f_j$. ($\mathbf{f}^p = (f_1, \ldots, f_j^p, \ldots, f_n)$). Because v_j is continuous in f_j, $\{v_j(\mathbf{x}, \mathbf{f}^p)\}_{p=1}^{\infty} \to v_j(\mathbf{x}, \mathbf{f})$. $\forall$ p,

$$\begin{aligned} & u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}^p), \ldots, f_j^p, \ldots, f_n) \\ = & u^*(f_1, \ldots, v_j(\mathbf{x}, \mathbf{f}^p), \ldots, f_n) \end{aligned}$$

and

$$\begin{aligned} & u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n) \\ = & u^*(f_1, \ldots, v_j(\mathbf{x}, \mathbf{f}), \ldots, f_n) \end{aligned}$$

By continuity of u^*,

$$\{u^*(f_1, \ldots, v_j(\mathbf{x}, \mathbf{f}^p), \ldots, f_n)\}_{p=1}^{\infty} \to u^*(f_1, \ldots, v_j(\mathbf{x}, \mathbf{f}), \ldots, f_p)$$

and therefore

$$\{u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}^p), \ldots, f_j^p, \ldots, f_n)\}_{p=1}^{\infty} \to u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n)$$

Therefore, $U = u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n)$ $\forall$ i satisfies Property 1. Property 2 is implied by the n-dimensional version of *Compromise/Attraction Monotonicity*. Properties 3 and 4 are still implied by Pairwise Transitivity and *Substitutability*, respectively.

To show the representation implies the axioms, it remains that the usual argument implies *Continuous Weak Order*, Property 4 implies *Substitutability*, Property 3 implies Pairwise Transitivity, Property 2 implies *Compromise/Attraction Monotonicity*, and strict increasing in the first n arguments implies *Simplicity*. This leaves *Frame Continuity* and *Pairwise Continuity*, and the two-dimensional proofs of these generalize easily to n dimensions.

A.2 Non-monotonic Attribute Preferences

In many applications, it is natural to have a good's desirability increase monotonically in an attribute. For example, all else equal, computers with more RAM, cars with more gas mileage, and televisions with better picture quality are all preferable. However, when considering the number of ports on a computer, or the color of a car, or the size of a television, it is not obvious that preferences over these attributes can be mapped to a monotonically increasing component.

However, attribute preferences can be established with the following properties, which are implied by the condition used in the paper:

1. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, if there exists $p \in \mathbb{R}^{n-1}$ such that $(x_i, p) \in C(\{(x_i, p), (y_i, p)\})$, then for each $p' \in \mathbb{R}^{n-1}$, $(x_i, p') \in C(\{(x_i, p'), (y_i, p')\})$.

2. "Attribute WARP (aWARP):" Given $S, S' \in \mathcal{S}$, $i \in \{1, \ldots, n\}$ such that $\forall\ s \in S, S'$, $s_{-i} = p_{-i}$ $(\forall\ -i \neq i)$; if $\mathbf{x}, \mathbf{y} \in S \cap S'$ and $\mathbf{x} \in C(S)$, then $\mathbf{y} \in C(S')$ implies $\mathbf{x} \in C(S')$.

3. $\{y \in \mathbb{R} | y \in C(\{(x, p), (y, p)\})\}$ and $\{z \in \mathbb{R} | x \in C(\{(x, p), (z, p)\})\}$ are both closed.

The first property establishes that the attributes are indeed distinct; if $x_i \succ y_i$ when they share some common set of other attribute values p, it remains so for any other set of attribute values p'. The second is a variant on WARP, applying only to cases where the differences exist only along one attribute. The final property ensures continuity.

Now, define $\mathbf{x} \succsim_i \mathbf{y}$ if and only if $(x_i, p) \in C(\{(x_i, p), (y_i, p)\})$ for some $p \in \mathbb{R}^{n-1}$. $\succsim_i$ is trivially complete. aWARP implies it is also transitive. Suppose $\mathbf{x} \succ_i \mathbf{y}$, $\mathbf{y} \succ_i \mathbf{z}$.

$$
\begin{aligned}
&(x_i, p) \in C\left(\{(x_i, p), (y_i, p)\}\right) \\
&(y_i, p) \in C\left(\{(y_i, p), (z_i, p)\}\right) \\
\Rightarrow &(y_i, p) \in C\left(\{(x_i, p), (y_i, p), (z_i, p)\}\right) \\
\Rightarrow &(x_i, p) \in C\left(\{(x_i, p), (y_i, p), (z_i, p)\}\right) \\
\Rightarrow &(x_i, p) \in C\left(\{(x_i, p), (z_i, p)\}\right) \\
&\quad \Rightarrow \mathbf{x} \succ_i \mathbf{z}
\end{aligned}
$$

Finally, because property 3 ensures $\succ_i$ is continuous, $\exists\ u_i$ representing $\succ_i\ \forall\ i$. We can now define a mapping $M : \mathbb{R}^n \to \mathbb{R}^n$ by $M(x) = (u_1(x), \ldots, u_n(x))$. This maps the goods to a space where their desirability is indeed increasing in the attributes, and the rest of the proof may proceed as written. These utility functions are not necessarily surjective, so the space of goods may not cover all of $\mathbb{R}^n$, but that is not a problem. This technique assumes that two goods $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x} \sim_i \mathbf{y}\ \forall\ i$ are treated identically.

Appendix B

Appendices to Extremist Politics and the Preference for Compromise

B.1 Proofs of Theorems

B.1.1 Multidimensional Citizen-Candidate without Compromise

Lemma 2. There does not exist an equilibrium in which a candidate x loses with certainty if $\forall\ i$ such that $\arg\max_{z\in K} u^i(e^z, s^z) = x$, $\arg\max_{z\in K/\{x\}} u^i(e^z, s^z) = y$ for some $y \in K/\{x\}$.

Proof of Lemma 1: Suppose for a contradiction that there does exist an equilibrium where candidate x loses with certainty, and the above condition is satisfied. The condition $\forall\ i$ such that $\arg\max_{z\in K} u^i(e^z, s^z) = x$, $\arg\max_{z\in K/\{x\}} u^i(e^z, s^z) = y$ requires that all voters who vote for candidate x agree on their second choice, candidate y. Note also that candidate x reaches their bliss point (in policy) at her own position, and thus $d(x,y) < d(x,z)\ \forall z \in K,\ z \neq x,\ y$.

In this equilibrium, candidate x obtains a payoff of $-c - d(x, k^*)$, where k^* is the ideal policy of the winner. If candidate x exits, candidate y obtains all of her previous voters. Thus, the new winner of the election is either candidate k^* who won when x had entered, or candidate y. Since y is preferred to k^* by x, the policy outcome of this election when x exits is weakly preferred by x. Since $c > 0$, x obtains a strictly higher payoff by exiting and therefore this equilibrium does not exist. ■

Corollary 1. In all two candidate equilibria, $v_x = v_y = \frac{1}{2}$.

Definition 1. An equilibrium is linear if there exists a function $f : e \to s$ such that for all $j \in K$, $s^j = ae^j + b$ for constants a, b.

Corollary 2. There does not exist a linear equilibrium with a sure loser z where $(e^z, s^z) << (e^y, s^y)$ $\forall y \in K$, $y \neq z$ or $(e^z, s^z) >> (e^y, s^y)$ $\forall y \in K$, $y \neq z$.

Note that Lemma 1 recovers Lemma 1 of Osborne and Slivinski (1996) as a special case (Corollary 2).

Proof of Proposition 1: Start with Proposition 1 (3). By Lemma 1, we know that there cannot be a sure loser in two candidate elections. Therefore, in any two-candidate equilibrium, vote shares must be $v_x = v_y = \frac{1}{2}$.

Furthermore, the candidates cannot be identical, i.e. $d(x, y) > 0$. Suppose, in equilibrium, the two candidates shared the same ideal policy. Then each candidate has a payoff of $\frac{1}{2}b - c$. If either candidate chose not to enter the race, her payoff would be 0, as the winner would share her ideal policy. Therefore, it must be the case that $b \geq 2c$. However, if the two candidates share the same position, then by continuity there exists another potential candidate with a distinct ideal policy who can enter and obtain arbitrarily close to half the votes. Let such a candidate be denoted by z. In equilibrium, she obtains payoff of $-d(x, z)$, but by entering would win with certainty and obtain payoff $b - c$. Since $b \geq 2c$, this potential candidate would not be best responding. Thus, $d(x, y) > 0$.

Since $d(x, y) > 0$ and $v_x = v_y$, each candidate obtains a payoff of $\frac{1}{2}b - \frac{1}{2}d(x, y) - c$ in equilibrium. If a candidate chose not to enter, she would obtain a payoff of $-d(x, y)$. Therefore, for candidates to choose to enter in equilibrium, we require $\frac{1}{2}b - \frac{1}{2}d(x, y) - c \geq -d(x, y)$, or $b \geq 2c - d(x, y)$. Therefore, for $b < 2c$, there exists a strictly positive lower bound on how differentiated candidates will be in all two candidate equilibria.

No potential candidate who is extreme with respect to party competition would like to enter. They obtain a strict subset of one equilibrium candidates' voters, and therefore lose the election. Furthermore, by entering they induce a victory by their least preferred equilibrium candidate. If $d(x,y) > d^*(F)$ then there exists a candidate who can enter and win the election. Since $b \geq 2c$

To prove Proposition 1(1), note that the requirements for all i such that $v_x(x,y,i) > v_y(x,y,i)$, $c \geq \frac{1}{2}[d(y,i) - d(x,i)]$ and for all i such that $v_y(x,y,i) > v_x(x,y,i)$, $c \geq \frac{1}{2}[d(x,i) - d(y,i)]$ guarantee that no candidate who can enter and influence the identity of the winner find it optimal to do so. If $b \geq 2c$, any distance $d(x,y) \in (0, d^*(F))$ (a nonempty interval) produces an equilibrium. If $b < 2c$, there exists an equilibrium if and only if $d^*(F) \geq 2c - b$. ■

Proof of Proposition 2: In order for x to be a sure winner, it must be the case that she obtains a larger vote share than any other candidate, and therefore $v_x > \max\{v_{L_1}; v_{L_2}\}$. Condition two requires that there does not exist a citizen who can enter and win with certainty, and would prefer to do so. If there exists a potential candidate z who could enter and win the election with certainty, then they would earn payoff $b-c$ by entering. In equilibrium, when choosing not to enter, z earns a payoff of $-d(x,z)$. Thus, if $b-c \leq -d(x,z)$, this candidate would choose not to enter. Therefore, for any citizen who can win with certainty by entering, we require $b \leq c - d(x,z)$. By continuity, if there exists a citizen who can enter and win with probability $p < 1$, then there exists a citizen who can enter and win with certainty. The payoff for a citizen j who can enter and win office with probability p is $pb - c - (1-p)d(x,j)$, and therefore the payoff restrictions in condition two also exclude entrants who are not sure winners.

Candidates L_1 and L_2 obtain payoff of $-c - d(L_1, x)$ and $-c - d(L_2, x)$ respectively. For L_1 to prefer to enter, it must be the case that she obtains a higher payoff by doing so. Suppose candidate x is still the winner when L_1 exits. Then, L_1 has no effect on the outcome by entering and thus would save c by exiting. Therefore, for this equilibrium to exist, it must be the case that L_2 wins if L_1 chooses not to enter. By exiting, L_1 obtains payoff $-d(L_1, L_2)$. Thus, for entry to be a best response, it must be the case that $c \le d(L_1, L_2) - d(L_1, x)$. The same conditions hold symmetrically for candidate L_2. These requirements imply conditions three and four.

Depending on the positions of each candidate, there may exist a citizen z who, when entering, extracts more votes from x than other candidates and can cause L_i to win the election. If such a citizen exists, he obtains $-d(x, z)$ in equilibrium, and $-c - d(L_i, z)$ by entering. Thus, to deter such a citizen from entering, it must be the case that $c \ge d(x, z) - d(L_i, z)$ for all L_i such that $v_{L_i} = \max_{k \in K \cup \{z\}}$.

Finally, x obtains payoff $b - c$ in equilibrium. Let L^* denote whichever of L_1 or L_2 would win the election in the event that x exited. Candidate x would therefore obtain a payoff of $-d(x, L^*)$ by exiting. Therefore, for this to be an equilibrium, it must be the case that $b \ge c - d(x, L^*)$. ■

B.1.2 Equilibria with Compromise Effects

Lemma 3. Assume that there exists two candidates with ideal policies $(0,0)$ and $(1,1)$. Assume that there exists two candidates x and y with ideal policies $(\frac{1}{2}-a, \frac{1}{2}-a)$ and $(\frac{1}{2}+a, \frac{1}{2}+a)$, $a \in [0, \frac{1}{2})$ respectively. $\forall i$ such that $e^i \ge \frac{1}{2}$ and $s^i \le \frac{1}{2}$, and $\forall i$ such that $e^i \le \frac{1}{2}$ and $s^i \ge \frac{1}{2}$, i votes for x or y.

Proof: Fix $e^i \le \frac{1}{2}$, $s^i \ge \frac{1}{2}$. The utility of voter i from voting for a candidate in this election is

$$U_i(e^x, s^x, \underline{e}^i, \underline{s}^i) = (1 - 2e^i + e^x)^{\frac{1}{2}} + (s^x)^{\frac{1}{2}} \qquad e^i > e^x$$

$$U_i(e^x, s^x, \underline{e}^i, \underline{s}^i) = (1 - e^x)^{\frac{1}{2}} + (s^x)^{\frac{1}{2}} \qquad e^i \le e^x$$

$$U_i(e^y, s^y, \underline{e}^i, \underline{s}^i) = (1 - e^y)^{\frac{1}{2}} + (2s^i - s^y)^{\frac{1}{2}} \qquad s^i < s^y$$

$$U_i(e^y, s^y, \underline{e}^i, \underline{s}^i) = (1 - e^y)^{\frac{1}{2}} + (s^y)^{\frac{1}{2}} \qquad s^i \ge s^y$$

$$U_i(0, 0, \underline{e}^i, \underline{s}^i) = (1 - 2e^i)^{\frac{1}{2}}$$

$$U_i(1, 1, \underline{e}^i, \underline{s}^i) = (2s^i - 1)^{\frac{1}{2}}$$

For all i with $e^i > e^x$, it is clear that $U_i(e^x, s^x, \underline{e}^i, \underline{s}^i) > U_i(0, 0, \underline{e}^i, \underline{s}^i)$. If $e^i \leq e^x$,

$$U_i(e^x, s^x, \underline{e}^i, \underline{s}^i) = (1 - (\frac{1}{2} - a))^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}}.$$

Note that $U_i(0, 0, \underline{e}^i, \underline{s}^i) = (1 - 2e^i)^{\frac{1}{2}} \leq 1$. If $U_i(e^x, s^x, \underline{e}^i, \underline{s}^i) \geq 1$, we can conclude that $x \succ L_1 \forall i$ such that $e^i \leq \frac{1}{2}$, $s^i \geq \frac{1}{2}$. This condition is satisfied if

$$\begin{aligned}
(1 - (\frac{1}{2} - a))^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}} &\geq 1 \\
(\frac{1}{2} + a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}} &\geq 1 \\
\frac{1}{2} + a + \frac{1}{2} - a + 2(\frac{1}{2} + a)^{\frac{1}{2}}(\frac{1}{2} - a)^{\frac{1}{2}} &\geq 1 \\
2(\frac{1}{2} + a)^{\frac{1}{2}}(\frac{1}{2} - a)^{\frac{1}{2}} &\geq 0
\end{aligned}$$

Since $(e^i, s^i) \in [0,1] \times [0,1]$, this condition always holds. For these voters, it remains to be shown that they will never choose to vote for L_2. Since $U_i(1, 1, \underline{e}^i, \underline{s}^i) = (2s^i - 1)^{\frac{1}{2}}$, the utility a voter receives from voting for L_2 is bounded above by 1. If $s^i \geq s^y$, $U_i(e^y, s^y, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} + a)^{\frac{1}{2}}$, and the above argument follows exactly. If $s^i < s^y$, $U_i(e^y, s^y, \underline{e}^i, \underline{s}^i)$ is bounded below by $(\frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - 1)^{\frac{1}{2}} = (\frac{1}{2} - a)^{\frac{1}{2}} + U_i(1, 1, \underline{e}^i, \underline{s}^i)$. Thus, all voters with ideal policy (e^i, s^i) satisfying $e^i \leq \frac{1}{2}$, $s^i \geq \frac{1}{2}$ vote for x or y. It remains to be shown that all voters with ideal policy (e^i, s^i) satisfying $e^i \geq \frac{1}{2}$, $s^i \leq \frac{1}{2}$ vote for x or y.

For all voters with ideal policy (e^i, s^i) satisfying $e^i \geq \frac{1}{2}$, $s^i \leq \frac{1}{2}$, the utility of voting for a candidate in this election is

$$U_i(e^x,s^x,\underline{e}^i,\underline{s}^i) = (e^x)^{\frac{1}{2}} + (1+s^x-2s^i)^{\frac{1}{2}} \qquad s^i > s^x$$
$$U_i(e^x,s^x,\underline{e}^i,\underline{s}^i) = (e^x)^{\frac{1}{2}} + (1-s^x)^{\frac{1}{2}} \qquad s^i \leq s^x$$
$$U_i(e^y,s^y,\underline{e}^i,\underline{s}^i) = (e^y)^{\frac{1}{2}} + (1-2s^i+s^y)^{\frac{1}{2}} \qquad e^i > e^y$$
$$U_i(e^y,s^y,\underline{e}^i,\underline{s}^i) = (2e^i-e^y)^{\frac{1}{2}} + (1-2s^i+s^y)^{\frac{1}{2}} \qquad e^i \leq e^y$$
$$U_i(0,0,\underline{e}^i,\underline{s}^i) = (1-2s^i)^{\frac{1}{2}}$$
$$U_i(1,1,\underline{e}^i,\underline{s}^i) = (2e^i-1)^{\frac{1}{2}}$$

If $s^x < s^i$, $U_i(e^x,s^x,\underline{e}^i,\underline{s}^i) = (e^x)^{\frac{1}{2}} + (1+s^x-2s^i)^{\frac{1}{2}}$ whereas $U_i(0,0,\underline{e}^i,\underline{s}^i) = (1-2s^i)^{\frac{1}{2}}$. Since e^x and $s^x \geq 0$, for these voters, $x \succ L_1$. Note that $U_i(0,0,\underline{e}^i,\underline{s}^i)$ is bounded above by 1. If $s^x \geq s^i$, $U_i(e^x,s^x,\underline{e}^i,\underline{s}^i) = (\frac{1}{2}+a)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}} \geq 1$ as shown above. Thus, $x \succ L_1$ for all voters with ideal policy (e^i,s^i) satisfying $e^i \geq \frac{1}{2}$, $s^i \leq \frac{1}{2}$.

If $e^i \geq e^y$, $U_i(e^y,s^y,\underline{e}^i,\underline{s}^i) = (\frac{1}{2}+a)^{\frac{1}{2}} + (\frac{3}{2}-2s^i+a)^{\frac{1}{2}}$, which is bounded below by $2(\frac{1}{2}+a)^{\frac{1}{2}}$. Similar to above, $U_i(1,1,\underline{e}^i,\underline{s}^i) = (2e^i-1)^{\frac{1}{2}}$ is bounded above by 1. Thus, for voters with $e^i \geq e^y$, $y \succ L_2$. If $e^i < e^y$, $U_i(e^y,s^y,\underline{e}^i,\underline{s}^i) = (2e^i-\frac{1}{2}-a)^{\frac{1}{2}} + (\frac{3}{2}-2s^i+a)^{\frac{1}{2}}$, which is bounded below by $(\frac{1}{2}-a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}}$. As shown above, for voters with $e^i < e^y$, $y \succ L_2$. Therefore, all voters with $e^i \geq \frac{1}{2}$, $s^i \leq \frac{1}{2}$ choose to vote for either x or y. ■

Lemma 4. Assume that there exists two candidates L_1 and L_2 with ideal policies $(0,0)$ and $(1,1)$ respectively. Assume that there exists two candidates x and y with ideal policies $(\frac{1}{2}-a,\frac{1}{2}-a)$ and $(\frac{1}{2}+a,\frac{1}{2}+a)$ respectively. L_1 obtains votes from all voters with $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ satisfying $s^i \leq \frac{1}{2} - \frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}} - (1-2e^i)^{\frac{1}{2}})^2$. L_2 obtains votes from all voters with $e^i > \frac{1}{2}$ and $s^i > \frac{1}{2}$ satisfying $s^i > \frac{1}{2} + \frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}} - (2e^i-1)^{\frac{1}{2}})^2$. Voters with ideal policies satisfying $e^i \in [\frac{1}{2}-a, \frac{1}{2}+a]$ vote for x if $s^i \leq 1-e^i$ and vote for y if $s^i > 1-e^i$. Voters with ideal policies satisfying $e^i < \frac{1}{2}-a$ vote for x if $s^i \leq \frac{1}{2}+a$ and they do not vote for L_1 and are indifferent between x and y if $s^i > \frac{1}{2}+a$. Voters satisfying $e^i > \frac{1}{2}+a$ vote for y if $s^i > \frac{1}{2}-a$ and they do not vote for L_2 and are indifferent between x and y if $s^i \leq \frac{1}{2}-a$.

Proof: **Voters with $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$:** Note that these voters will either vote for L_1

or x. Their frame $(\underline{e}^i, \underline{s}^i)$ is given by $(1,1)$; for a fixed frame, voters utility of voting for a candidate is decreasing in the difference in ideal policies. Thus, these voters strictly prefer both L_1 and x to y and L_2.

Voters with $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ have utility $(1-2e^i)^{\frac{1}{2}} + (1-2s^i)^{\frac{1}{2}}$ of voting for L_1. This set of voters differs in their utility of voting for x by whether they have $e^i \geq \frac{1}{2} - a$, $s^i \geq \frac{1}{2} - a$, both or neither. Their utility is given by

$$U_i(\frac{1}{2}-a, \frac{1}{2}-a, 1, 1) = \begin{cases} 2(\frac{1}{2}+a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}-a,\ s^i \leq \frac{1}{2}-a \\ (\frac{1}{2}+a)^{\frac{1}{2}} + (\frac{3}{2}-2s^i-a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}-a,\ s^i > \frac{1}{2}-a \\ (\frac{3}{2}-2e^i-a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}} & e^i > \frac{1}{2}-a,\ s^i \leq \frac{1}{2}-a \\ (\frac{3}{2}-2e^i-a)^{\frac{1}{2}} + (\frac{3}{2}-2s^i-a)^{\frac{1}{2}} & e^i > \frac{1}{2}-a,\ s^i > \frac{1}{2}-a \end{cases}$$

Since $\frac{3}{2} - a > 1$, for all voters with $e^i > \frac{1}{2} - a$ and $s^i > \frac{1}{2} - a$, x is preferred to L_1. For voters with $e^i \leq \frac{1}{2} - a$ and $s^i > \frac{1}{2} - a$, their utility of voting for x is $(\frac{1}{2}+a)^{\frac{1}{2}} + (\frac{3}{2}-2s^i-a)^{\frac{1}{2}}$, whereas their utility of voting for L_1 is $(1-2e^i)^{\frac{1}{2}} + (1-2s^i)^{\frac{1}{2}}$. Their utility of voting for L_1 is bounded above by $1 + (1-2s^i)^{\frac{1}{2}}$, whereas the utility of voting for x is constant in e^i. Consider the voter with $s^i = \frac{1}{2}$. His utility of voting for x is $(\frac{1}{2}+a)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}} > 1$ for all $a < \frac{1}{2}$, whereas his utility of voting for L_1 is 1. For all values of $a \in [0, \frac{1}{2})$, this voter strictly prefers x. Now consider the voter with $s^i = \frac{1}{2} - a$. His utility of voting for x is $2(\frac{1}{2}+a)^{\frac{1}{2}}$, and his utility of voting for L_1 is $1 + (2a)^{\frac{1}{2}}$. If $a = 0$, he strictly prefers voting for x to voting for L_1. If $a = \frac{1}{2}$, he is indifferent between x and L_1. By continuity and monotonicity of $U(.)$ in a, this voter strictly prefers x to L_1 for all $a \in [0, \frac{1}{2})$. Thus, all voters with $e^i \leq \frac{1}{2} - a$ and $s^i > \frac{1}{2} - a$ strictly prefer x to L_1. By symmetry, this is also true for voters with $e^i > \frac{1}{2} - a$ and $s^i \leq \frac{1}{2} - a$. Finally, voters with $e^i \leq \frac{1}{2} - a$ and $s^i \leq \frac{1}{2} - a$ have utility $2(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for x and utility of $(1-2e^i)^{\frac{1}{2}} + (1-2s^i)^{\frac{1}{2}}$ of voting for L_1. The set of indifferent voters is given by $s^i = \frac{1}{2} - \frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}} - (1-2e^i)^{\frac{1}{2}})^2$. Since the utility of voting for L_1 is decreasing in both e^i and s^i, and the utility of voting for x is constant with respect to e^i and s^i for these voters, voters with $s^i \leq \frac{1}{2} - \frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}} - (1-2e^i)^{\frac{1}{2}})^2$ vote for L_1, whereas voters with $s^i > \frac{1}{2} - \frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}} - (1-2e^i)^{\frac{1}{2}})^2$ vote for x.

Voters with $e^i \geq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$: As above, the fact that $U(.)$ is monotonically decreasing in difference in ideal policies from a candidate implies that these voters vote for L_2 or y.

These voters have utility $U(1,1,0,0) = (2e^i-1)^{\frac{1}{2}} + (2s^i-1)^{\frac{1}{2}}$ of voting for L_2 and

$$U(\frac{1}{2}+a, \frac{1}{2}+a, 0, 0) = \begin{cases} (2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}+a,\ s^i \leq \frac{1}{2}+a \\ (2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}+a,\ s^i > \frac{1}{2}+a \\ (\frac{1}{2}+a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} & e^i > \frac{1}{2}+a,\ s^i \leq \frac{1}{2}+a \\ 2(\frac{1}{2}+a)^{\frac{1}{2}} & e^i > \frac{1}{2}+a,\ s^i > \frac{1}{2}+a \end{cases}$$

of voting for y.

Voters with $e^i \leq \frac{1}{2}+a$ and $s^i \leq \frac{1}{2}+a$ have utility $(2e^i-\frac{1}{2}-a)^{\frac{1}{2}}+(2s^i-\frac{1}{2}-a)^{\frac{1}{2}}$ of voting for y and $(2e^i-1)^{\frac{1}{2}}+(2s^i-1)^{\frac{1}{2}}$ of voting for L_2. Since $a \in [0, \frac{1}{2})$, these voters all prefer y to L_2. Voters with $e^i \leq \frac{1}{2}+a$ and $s^i > \frac{1}{2}+a$ have utility $(2e^i-1)^{\frac{1}{2}}+(2s^i-1)^{\frac{1}{2}}$ of voting for L_2 and $(2e^i-\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for y. Note that $(2e^i-1)^{\frac{1}{2}}+(2s^i-1)^{\frac{1}{2}}$ is bounded above by $(2e^i-1)^{\frac{1}{2}}+1$. Suppose $e^i = \frac{1}{2}$; this voter has utility $(\frac{1}{2}+a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}}$ of voting for y which is strictly larger than his utility of voting for L_2 which is 1. If $e^i = \frac{1}{2}+a$, this voter has utility $2(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for y and $1+(2a)^{\frac{1}{2}}$ of voting for L_2. If $a = 0$, the utility of voting for y is strictly larger than that of voting for L_2. If $a = \frac{1}{2}$, he obtains utility 2 of voting for both y and L_2 and is indifferent. Since $a \in [0, \frac{1}{2})$ and continuity of $U(.)$ in a, all voters with $e^i = \frac{1}{2}+a$ and $s^i > \frac{1}{2}+a$ prefer y to L_2. By continuity and monotonicity of $U(.)$ in e^i for these voters, all voters with $e^i \leq \frac{1}{2}+a$ and $s^i > \frac{1}{2}+a$ prefer y to L_2. Symmetrically, voters with $e^i > \frac{1}{2}+a$ and $s^i \leq \frac{1}{2}+a$ prefer y to L_2. Voters with $e^i > \frac{1}{2}+a$ and $s^i > \frac{1}{2}+a$ have utility $2(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for y

and $(2e^i-1)^{\frac{1}{2}}+(2s^i-1)^{\frac{1}{2}}$ of voting for L_2. The set of indifferent voters is given by $s^i=\frac{1}{2}+\frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}}-(2e^i-1)^{\frac{1}{2}})^2$. Since these voters' utility of voting for y is constant and increasing in s^i and e^i for L_2, voters with $s^i>\frac{1}{2}+\frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}}-(2e^i-1)^{\frac{1}{2}})^2$ vote for L_2, whereas those with $s^i<\frac{1}{2}+\frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}}-(2e^i-1)^{\frac{1}{2}})^2$ vote for y.

Voters with $e^i>\frac{1}{2}$ and $s^i\leq\frac{1}{2}$: By Lemma 3, no voters with $e^i>\frac{1}{2}$ and $s^i\leq\frac{1}{2}$ vote for L_1 or L_2. These voters have utility of voting for x and y given by

$$U(\frac{1}{2}-a,\frac{1}{2}-a,0,1)=\begin{cases}(\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{3}{2}-2s^i-a)^{\frac{1}{2}} & s^i>\frac{1}{2}-a\\ (\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}+a)^{\frac{1}{2}} & s^i\leq\frac{1}{2}-a\end{cases}$$

and

$$U(\frac{1}{2}+a,\frac{1}{2}+a,0,1)=\begin{cases}(\frac{1}{2}+a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}} & e^i>\frac{1}{2}+a\\ (2e^i-\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}} & e^i\leq\frac{1}{2}+a\end{cases}$$

respectively. Voters with (e^i,s^i) satisfying $e^i>\frac{1}{2}+a$ and $s^i\leq\frac{1}{2}-a$ have utility of $(\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for each candidate, and are thus indifferent. Voters with $e^i>\frac{1}{2}+a$ and $s^i>\frac{1}{2}-a$ have utility $(\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for y and utility $(\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{3}{2}-2s^i-a)^{\frac{1}{2}}$ of voting for x. Since their utility of voting for x is bounded above by their utility of voting for y, these voters therefore vote for y. Symmetrically, voters with $e^i\leq\frac{1}{2}+a$ and $s^i\leq\frac{1}{2}-a$ have strictly larger utility of voting for x. Voters with $e^i\leq\frac{1}{2}+a$ and $s^i>\frac{1}{2}-a$ have utility $(\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{3}{2}-2s^i-a)^{\frac{1}{2}}$ of voting for x and $(2e^i-\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}}$ of voting for y. The set of indifferent voters is given by $(\frac{3}{2}-2s^i-a)^{\frac{1}{2}}=(2e^i-\frac{1}{2}-a)^{\frac{1}{2}}$, which simplifies to $s^i=1-e^i$. Since the utility of voting for x is decreasing in s^i and the utility of voting for y is increasing in e^i, voters in this subpopulation with $s^i>1-e^i$ vote for y and those with $s^i<1-e^i$ vote for x.

Voters with $e^i\leq\frac{1}{2}$ and $s^i>\frac{1}{2}$: By Lemma 3, no voters with $e^i\leq\frac{1}{2}$ and $s^i>\frac{1}{2}$

vote for L_1 or L_2. These voters have utility of voting for x and y given by

$$U(\frac{1}{2}-a,\frac{1}{2}-a,1,0)=\begin{cases}(\frac{1}{2}+a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}-a\\(\frac{3}{2}-a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}} & e^i > \frac{1}{2}-a\end{cases}$$

and

$$U(\frac{1}{2}+a,\frac{1}{2}+a,1,0)=\begin{cases}(\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}+a)^{\frac{1}{2}} & s^i > \frac{1}{2}+a\\(\frac{1}{2}-a)^{\frac{1}{2}}+(2s^i-\frac{1}{2}-a)^{\frac{1}{2}} & s^i \leq \frac{1}{2}+a\end{cases}$$

respectively. For voters with $e^i \leq \frac{1}{2}-a$ and $s^i > \frac{1}{2}+a$, the utility of voting for x and y is the same, and they are indifferent. Voters with $e^i \leq \frac{1}{2}-a$ and $s^i \leq \frac{1}{2}+a$ have utility $(\frac{1}{2}+a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}}$ of voting for x and $(\frac{1}{2}-a)^{\frac{1}{2}}+(2s^i-\frac{1}{2}-a)^{\frac{1}{2}}$ of voting for y. Since $(2s^i-\frac{1}{2}-a)^{\frac{1}{2}}$ is bounded above by $(\frac{1}{2}+a)^{\frac{1}{2}}$ for this subpopulation of voters, they prefer x to y. Symmetrically, voters with $e^i > \frac{1}{2}-a$ and $s^i > \frac{1}{2}-a$ prefer y to x. Voters with $e^i > \frac{1}{2}-a$ and $s^i \leq \frac{1}{2}+a$ have utility $(\frac{3}{2}-a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}}$ of voting for x and $(\frac{1}{2}-a)^{\frac{1}{2}}+(2s^i-\frac{1}{2}-a)^{\frac{1}{2}}$ of voting for y. The set of indifferent voters is given by $(\frac{3}{2}-a-2e^i)^{\frac{1}{2}}=(2s^i-\frac{1}{2}-a)^{\frac{1}{2}}$, which reduces to $1-e^i=s^i$. Since the utility of voting for x is decreasing in e^i for these voters and the utility of voting for y is increasing in s^i, voters with $s^i > 1-e^i$ vote for y and those with $s^i \leq 1-e^i$ vote for x. ■

Lemma 5. Assume there exists candidates L_1 and L_2 with $(e^i,s^i)=(0,0)$ and $(1,1)$ respectively. Assume there exists a candidate y with $(e^i,s^i)=(\frac{1}{2}+a,\frac{1}{2}+a)$. L_2 receives votes from all voters with $e^i > \frac{1}{2}$ and $s^i > \frac{1}{2}$ satisfying $s^i > \frac{1}{2}+\frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}}-(2e^i-1)^{\frac{1}{2}})^2$. L_1 receives votes from voters with $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ satisfying $s^i \leq \frac{1}{2}-\frac{1}{2}(2(\frac{1}{2}-a)^{\frac{1}{2}}-(1-2e^i)^{\frac{1}{2}})^2$, voters with $e^i \leq \frac{1}{2}$ and $\frac{1}{2} < s^i \leq \frac{1}{2}+a$ satisfying $s^i \leq \frac{1}{4}+\frac{a}{2}+\frac{1}{2}((1-2e^i)^{\frac{1}{2}}-(\frac{1}{2}-a)^{\frac{1}{2}})^2$, and voters with $\frac{1}{2} < e^i \leq \frac{1}{2}+a$ and $s^i \leq \frac{1}{2}$ satisfying $s^i \leq \frac{1}{2}-\frac{1}{2}((2e^i-\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}})^2$. The remaining voters vote for y.

Proof: First, note that the frame is the same for all voters as in the previous two lemmas. By Lemma 3, y is preferred to L_2 for all voters with $e^i \leq \frac{1}{2}$ and $s^i > \frac{1}{2}$ and all voters $e^i > \frac{1}{2}$ and $s^i \leq \frac{1}{2}$. By Lemma 4, L_2 receives votes from all voters with $e^i > \frac{1}{2}$

and $s^i > \frac{1}{2}$ satisfying $s^i > \frac{1}{2} + \frac{1}{2}(2(\frac{1}{2}+a)^{\frac{1}{2}} - (2e^i - 1)^{\frac{1}{2}})^2$.

Voters with $e^i \leq \frac{1}{2}$ and $s^i > \frac{1}{2}$. These voters have utility of voting for y given by

$$U(e^y, s^y, 1, 0) = \begin{cases} (\frac{1}{2}-a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}} & s^i > \frac{1}{2}+a \\ (\frac{1}{2}-a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} & s^i \leq \frac{1}{2}+a \end{cases}$$

and utility of voting for L_1 given by

$$U(0,0,1,0) = (1-2e^i)^{\frac{1}{2}}.$$

Since $(\frac{1}{2}-a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}} > 1$ for all $a < \frac{1}{2}$, and $(1-2e^i)^{\frac{1}{2}} \leq 1$, all voters with $e^i \leq \frac{1}{2}$ and $s^i > \frac{1}{2}+a$ vote for y. The set of indifferent voters between L_1 and y is given by voters with $s^i \leq \frac{1}{2}+a$ satisfying $s^i = \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((1-2e^i)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$. Since the utility of voting for y is monotonically increasing in s^i and the utility of voting for L_1 is monotonically decreasing in e^i, voters with $s^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((1-2e^i)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for L_1 and those with $s^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((1-2e^i)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for y.

Voters with $e^i > \frac{1}{2}$ and $s^i \leq \frac{1}{2}$. These voters have utility of voting for y given by

$$U(e^y, s^y, 0, 1) = \begin{cases} (\frac{1}{2}+a)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}} & e^i > \frac{1}{2}+a \\ (2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}+a \end{cases}$$

and utility of voting for L_1 given by

$$U(0,0,0,1) = (1-2s^i)^{\frac{1}{2}}.$$

Since $(\frac{1}{2}-a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}} > 1$ for all $a < \frac{1}{2}$, and $(1-2s^i)^{\frac{1}{2}} \leq 1$, all voters with $e^i > \frac{1}{2}+a$ vote for y. For voters with $e^i \leq \frac{1}{2}+a$, the set of indifferent voters between y and L_1 is given by $s^i = \frac{1}{2} - \frac{1}{2}((2e^i - \frac{1}{2} - a)^{\frac{1}{2}} - (1-2e^i)^{\frac{1}{2}})^2$. Since the utility of voting for y is monotonically increasing in e^i and the utility of voting for L_1 is monotonically decreasing in s^i, voters with $s^i \leq \frac{1}{2} - \frac{1}{2}((2e^i - \frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} - a)^{\frac{1}{2}})^2$ vote for L_1, and those with $s^i > \frac{1}{2} - \frac{1}{2}((2e^i - \frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} - a)^{\frac{1}{2}})^2$ vote for y.

Voters with $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$. These voters have utility of voting for y given by

$$U(e^y, s^y, 1, 1) = 2(\frac{1}{2} - a)^{\frac{1}{2}}$$

and utility of voting for L_1 given by

$$U(0, 0, 1, 1) = (1 - 2e^i)^{\frac{1}{2}} + (1 - 2s^i)^{\frac{1}{2}}.$$

The set voters who are indifferent between y and L_1 is given by $s^i = \frac{1}{2} - \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (1 - 2e^i)^{\frac{1}{2}})^2$. Since the utility of voting for y is constant with respect to e^i and s^i, and the utility of voting for L_1 is monotonically decreasing in both e^i and s^i, voters with $s^i \leq \frac{1}{2} - \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (1 - 2e^i)^{\frac{1}{2}})^2$ vote for L_1 and voters with $s^i > \frac{1}{2} - \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (1 - 2e^i)^{\frac{1}{2}})^2$ vote for y. ■

Lemma 6. Assume there exists candidates L_1 and L_2 with $(e^i, s^i) = (0, 0)$ and $(1, 1)$ respectively. Assume there exists a candidate x with $(e^i, s^i) = (\frac{1}{2} - a, \frac{1}{2} - a)$. L_1 receives votes from all voters with $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ satisfying $s^i \leq \frac{1}{2} - \frac{1}{2}(2(\frac{1}{2} + a)^{\frac{1}{2}} - (1 - 2e^i)^{\frac{1}{2}})^2$. L_2 receives votes from all voters with $\frac{1}{2} - a < e^i \leq \frac{1}{2}$ and $s^i > \frac{1}{2}$ satisfying $si > \frac{1}{2} + \frac{1}{2}((\frac{3}{2} - a - 2e^i)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}})^2$, voters with $e^i > \frac{1}{2}$ and $\frac{1}{2} - a < s^i \leq \frac{1}{2}$ satisfying $si > \frac{3}{4} - \frac{a}{2} - \frac{1}{2}((2e^i - 1)^{\frac{1}{2}} - (\frac{1}{2} - a)^{\frac{1}{2}})^2$, and voters with $e^i > \frac{1}{2}$ and $s^i > \frac{1}{2}$ satisfying $s^i > \frac{1}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (2e^i - 1)^{\frac{1}{2}})^2$. The remaining voters vote for x.

Proof: Note that the frame is the same for all voters as in the preceding lemmas. By Lemma 3, x is preferred to L_1 for all voters with $e^i \leq \frac{1}{2}$ and $s^i > \frac{1}{2}$ and all voters with $e^i > \frac{1}{2}$ and $s^i \leq \frac{1}{2}$. By Lemma 4, L_1 receives votes from all voters with $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ satisfying $s^i \leq \frac{1}{2} - \frac{1}{2}(2(\frac{1}{2} + a)^{\frac{1}{2}} - (1 - 2e^i)^{\frac{1}{2}})^2$.

Voters with $e^i > \frac{1}{2}$ and $s^i \leq \frac{1}{2}$: These voters have utility of voting for x given by

$$U(\frac{1}{2}-a,\frac{1}{2}-a,0,1)=\begin{cases}(\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{3}{2}-2s^i-a)^{\frac{1}{2}} & s^i > \frac{1}{2}-a\\ (\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}+a)^{\frac{1}{2}} & s^i \leq \frac{1}{2}-a\end{cases}$$

and utility of voting for L_2 given by

$$U(1,1,0,1)=(2e^i-1)^{\frac{1}{2}}.$$

Since $2e^i - 1 \leq 1$ for all e^i, all voters with $s^i \leq \frac{1}{2} - a$ prefer x to L_2. The set of voters with $s^i > \frac{1}{2}$ indifferent between voting for x and L_2 is given by $s^i = \frac{3}{4} - \frac{a}{2} - \frac{1}{2}((2e^i-1)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$. Since the utility of voting for x is monotonically decreasing in s^i and the utility of voting for L_2 is monotonically increasing in e^i, voters with $s^i > \frac{3}{4}-\frac{a}{2}-\frac{1}{2}((2e^i-1)^{\frac{1}{2}}-(\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for L_2 and those with $s^i \leq \frac{3}{4}-\frac{a}{2}-\frac{1}{2}((2e^i-1)^{\frac{1}{2}}-(\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for x.

Voters with $e^i \leq \frac{1}{2}$ and $s^i > \frac{1}{2}$: These voters have utility of voting for x given by

$$U(\frac{1}{2}-a,\frac{1}{2}-a,0,1)=\begin{cases}(\frac{3}{2}-a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}} & e^i > \frac{1}{2}-a\\ (\frac{1}{2}+a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}-a\end{cases}$$

and utility of voting for L_2 given by

$$U(1,1,0,1)=(2s^i-1)^{\frac{1}{2}}.$$

All voters with $e^i \leq \frac{1}{2}-a$ prefer voting for x to L_2. The set of voters with $e^i > \frac{1}{2}-a$ who are indifferent between x and L_2 is given by $s^i = \frac{1}{2}+\frac{1}{2}((\frac{3}{2}-a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}})^2$. Since the utility of voting for x is monotonically decreasing in e^i and the utility of voting for L_2 is monotonically increasing in s^i, voters with $s^i > \frac{1}{2}+\frac{1}{2}((\frac{3}{2}-a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for L_2 and voters with $s^i \leq \frac{1}{2}+\frac{1}{2}((\frac{3}{2}-a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for x.

Voters with $e^i > \frac{1}{2}$ and $s^i > \frac{1}{2}$: These voters have utility of voting for x given by

$$U(\frac{1}{2}-a,\frac{1}{2}-a,0,0) = 2(\frac{1}{2}-a)^{\frac{1}{2}}$$

and utility of voting for L_2 given by

$$U(1,1,0,0) = (2e^i-1)^{\frac{1}{2}} + (2s^i-1)^{\frac{1}{2}}.$$

The set of voters indifferent between x and L_2 is given by $s^i = \frac{1}{2}+\frac{1}{2}(2(\frac{1}{2}-a)^{\frac{1}{2}} - (2e^i-1)^{\frac{1}{2}})^2$. Since the utility of voting for x is constant and the utility of voting for L_2 is monotonically increasing in both e^i and s^i, voters with $s^i > \frac{1}{2}+\frac{1}{2}(2(\frac{1}{2}-a)^{\frac{1}{2}}-(2e^i-1)^{\frac{1}{2}})^2$ vote for L_2 and voters with $s^i \le \frac{1}{2}+\frac{1}{2}(2(\frac{1}{2}-a)^{\frac{1}{2}}-(2e^i-1)^{\frac{1}{2}})^2$ vote for x. ∎

Lemma 7. Assume there exist candidates L_1, x, and y with ideal policies $(0,0)$, $(\frac{1}{2}-a,\frac{1}{2}-a)$, and $(\frac{1}{2}+a,\frac{1}{2}+a)$ respectively.

1. If $a \le \frac{1}{6}$, L_1 obtains votes from all voters with $e^i \le \frac{1}{4}+\frac{a}{2}$ and $s^i \le \frac{1}{4}+\frac{a}{2}$ satisfying $s^i \le \frac{1}{4}+\frac{a}{2}-\frac{1}{2}(2(2a)^{\frac{1}{2}}-(\frac{1}{2}+a-2e^i)^{\frac{1}{2}})^2$, voters with $e^i \le \frac{1}{4}+\frac{a}{2}$ and $s^i > \frac{1}{4}+\frac{a}{2}$ satisfying $s^i \le \frac{1}{4}-\frac{a}{2}+\frac{1}{2}((\frac{1}{2}+a-2e^i)^{\frac{1}{2}}-(2a)^{\frac{1}{2}})^2$, and voters with $e^i > \frac{1}{4}+\frac{a}{2}$ and $s^i \le \frac{1}{4}+\frac{a}{2}$ satisfying $s^i \le \frac{1}{4}+\frac{a}{2}-\frac{1}{2}((2e^i-\frac{1}{2}+a)^{\frac{1}{2}}-(2a)^{\frac{1}{2}})^2$.

   Candidate y obtains votes from all voters with $e^i > \frac{1}{2}+a$ or $s^i > \frac{1}{2}+a$. Candidate y additionally obtains votes from voters with $\frac{1}{4}+\frac{a}{2} < e^i \le \frac{1}{2}-a$ and $s^i > \frac{1}{2}-a$ satisfying $s^i > \frac{1}{4}+\frac{a}{2}+\frac{1}{2}((2e^i-\frac{1}{2}+a)^{\frac{1}{2}}-(2e^i-\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}})^2$, voters with $\frac{1}{4}+\frac{a}{2} < s^i \le \frac{1}{2}-a$ and $e^i > \frac{1}{2}-a$ satisfying $e^i > \frac{1}{4}+\frac{a}{2}+\frac{1}{2}((2s^i-\frac{1}{2}+a)^{\frac{1}{2}} - (2s^i-\frac{1}{2}-a)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}})^2$, and voters with $e^i > \frac{1}{2}-a$ and $s^i > \frac{1}{2}-a$ satisfying $s^i > \frac{1}{4}+\frac{a}{2}+\frac{1}{2}(2(\frac{1}{2}-a)^{\frac{1}{2}}-(2e^i-\frac{1}{2}-a)^{\frac{1}{2}})^2$.

   The remaining voters vote for x.

2. If $a > \frac{1}{6}$, L_1 obtains votes from all voters with $e^i \le \frac{1}{2}-a$ and $s^i \le \frac{1}{2}-a$ satisfying $s^i \le \frac{1}{4}+\frac{a}{2}-\frac{1}{2}(2(2a)^{\frac{1}{2}}-(\frac{1}{2}+a-2e^i)^{\frac{1}{2}})^2$, voters with $e^i \le \frac{1}{2}-a$ and $\frac{1}{2}-a < s^i \le \frac{1}{4}+\frac{a}{2}$ satisfying $e^i \le \frac{1}{4}+\frac{a}{2}-\frac{1}{2}(2a)^{\frac{1}{2}}+(1-2s^i)^{\frac{1}{2}}-(\frac{1}{2}+a-2s^i)^{\frac{1}{2}})^2$, and voters with $\frac{1}{2}-a < e^i \le \frac{1}{4}+\frac{a}{2}$ satisfying $s^i \le \frac{1}{4}+\frac{a}{2}-\frac{1}{2}((2a)^{\frac{1}{2}}+(1-2e^i)^{\frac{1}{2}}-(\frac{1}{2}+a-2e^i)^{\frac{1}{2}})^2$.

   Candidate y obtains votes from all voters with $e^i > \frac{1}{2}+a$ and $s^i > \frac{1}{2}+a$. Candidate y also obtains votes from voters with $e^i > \frac{1}{2}+a$ and $\frac{1}{4}+\frac{a}{2} < s^i \le \frac{1}{2}+a$ satisfying

$s^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} + a)^{\frac{1}{2}})^2$. Voters with $s^i > \frac{1}{2} + a$ and $\frac{1}{4} + \frac{a}{2} < e^i \le \frac{1}{2} + a$ vote for y if $e^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} + a)^{\frac{1}{2}})^2$. Voters with $\frac{1}{4} + \frac{a}{2} < e^i \le \frac{1}{2} + a$ and $\frac{1}{4} + \frac{a}{2} < s^i \le \frac{1}{2} + a$ vote for y if $s^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2$. Voters with $e^i > \frac{1}{2} + a$ and $\frac{1}{2} - a < s^i \le \frac{1}{4} + \frac{a}{2}$ vote for y if $s^i > \frac{1}{2} - \frac{1}{2}((\frac{1}{2} + a)^{\frac{1}{2}} - (\frac{1}{2} - a)^{\frac{1}{2}})^2$. Voters with $\frac{1}{4} + \frac{a}{2} < e^i \le \frac{1}{2} + a$ and $\frac{1}{2} - a < s^i \le \frac{1}{4} + \frac{a}{2}$ vote for y if $s^i > \frac{1}{2} - \frac{1}{2}((2e^i - \frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} - a)^{\frac{1}{2}})^2$. Voters with $\frac{1}{2} - a < e^i \le \frac{1}{4} + \frac{a}{2}$ and $\frac{1}{4} + \frac{a}{2} < s^i \le \frac{1}{2} + a$ vote for y if $s^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((1 - 2e^i)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}})^2$. Voters with $\frac{1}{2} - a < e^i \le \frac{1}{4} + \frac{a}{2}$ and $\frac{1}{2} + a < s^i$ vote for y if $e^i > \frac{1}{2} - \frac{1}{2}((\frac{1}{2} + a)^{\frac{1}{2}} - (\frac{1}{2} - a)^{\frac{1}{2}})^2$.

Proof: Note that $\underline{e}^i = \max_{j \in X} |e^j - e^i|$ and $\underline{s}^i = \max_{j \in X} |s^j - s^i|$. Candidates have positions $(0,0)$, $(\frac{1}{2} - a, \frac{1}{2} - a)$ and $(\frac{1}{2} + a, \frac{1}{2} + a)$. Thus, all citizens with $e^i \le \frac{1}{4} + \frac{a}{2}$ have $\underline{e}^i = \frac{1}{2} + a$ and citizens with $e^i > \frac{1}{4} + \frac{a}{2}$ have $\underline{e}^i = 0$. Similarly, citizens with $s^i \le \frac{1}{4} + \frac{a}{2}$ have $\underline{s}^i = \frac{1}{2} + a$ and citizens with $s^i > \frac{1}{4} + \frac{a}{2}$ have $\underline{s}^i = 0$. The frames divide the unit square into four subpopulations by e^i and s^i. Candidate x lies in the group of citizens with $\underline{e}^i = \underline{s}^i = 0$ if $\frac{1}{4} + \frac{a}{2} \le \frac{1}{2} - a$, or $a \le \frac{1}{6}$. Alternatively, if $a > \frac{1}{6}$, x lies in the group of citizens with $\underline{e}^i = \underline{s}^i = \frac{1}{2} + a$. Candidate y always lies in the group of citizens with $\underline{e}^i = \underline{s}^i = 0$ and candidate L_1 is always in the group of citizens with $\underline{e}^i = \underline{s}^i = \frac{1}{2} + a$.

Consider the case with $a \le \frac{1}{6}$:

Voters with $e^i \le \frac{1}{4} + \frac{a}{2}$ and $s^i \le \frac{1}{4} + \frac{a}{2}$: These voters have utility of voting for x given by

$$U(\frac{1}{2} - a, \frac{1}{2} - a, \frac{1}{2} + a, \frac{1}{2} + a) = 2(2a)^{\frac{1}{2}}$$

and utility of voting for y given by

$$U(\frac{1}{2} + a, \frac{1}{2} + a, \frac{1}{2} + a, \frac{1}{2} + a) = 0.$$

Clearly, none of these voters will vote for y. Their utility of voting for L_1 is given by

$$U(0, 0, \frac{1}{2} + a, \frac{1}{2} + a) = (\frac{1}{2} + a - 2e^i)^{\frac{1}{2}} + (\frac{1}{2} + a - 2s^i)^{\frac{1}{2}}.$$

The set of voters who are indifferent between x and L_1 is therefore given by

$$2(2a)^{\frac{1}{2}} = (\frac{1}{2} + a - 2e^i)^{\frac{1}{2}} + (\frac{1}{2} + a - 2s^i)^{\frac{1}{2}}$$

which reduces to $s^i = \frac{1}{4} + \frac{a}{2} - \frac{1}{2}(2(2a)^{\frac{1}{2}} - (\frac{1}{2} + a - 2e^i)^{\frac{1}{2}})^2$. Since these voters all have the same utility of voting for x whereas their utility of voting for L_1 is strictly decreasing in e^i and s^i, voters with $s^i \leq \frac{1}{4} + \frac{a}{2} - \frac{1}{2}(2(2a)^{\frac{1}{2}} - (\frac{1}{2} + a - 2e^i)^{\frac{1}{2}})^2$ vote for L_1 and those with $si > \frac{1}{4} + \frac{a}{2} - \frac{1}{2}(2(2a)^{\frac{1}{2}} - (\frac{1}{2} + a - 2e^i)^{\frac{1}{2}})^2$ vote for x.

Voters with $e^i \leq \frac{1}{4} + \frac{a}{2}$ and $s^i > \frac{1}{4} + \frac{a}{2}$: These voters have utility of voting for L_1 given by

$$U(0, 0, \frac{1}{2} + a, 0) = (\frac{1}{2} + a - 2e^i)^{\frac{1}{2}}.$$

Their utility of voting for x is given by

$$U(\frac{1}{2} - a, \frac{1}{2} - a, \frac{1}{2} + a, 0) = \begin{cases} (2a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} + a)^{\frac{1}{2}} & s^i \leq \frac{1}{2} - a \\ (2a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}} & s^i > \frac{1}{2} - a \end{cases}$$

and their utility of voting for y is given by

$$U(\frac{1}{2} + a, \frac{1}{2} + a, \frac{1}{2} + a, 0) = \begin{cases} (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} & s^i \leq \frac{1}{2} + a \\ (\frac{1}{2} + a)^{\frac{1}{2}} & s^i > \frac{1}{2} + a \end{cases}$$

and thus all voters with $s^i \leq \frac{1}{2} - a$ prefer x to y. Voters with $s^i > \frac{1}{2} + a$ prefer x to y if $(\frac{1}{2} + a)^{\frac{1}{2}} < (2a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}}$. This condition holds by concavity, and thus these voters prefer x to y. By continuity and monotonicity, voters with $s^i \in (\frac{1}{2} - a, \frac{1}{2} + a]$ prefer x to y as well.

Since the utility of voting for L_1 is bounded above by $(\frac{1}{2} + a)^{\frac{1}{2}}$, all voters with $s^i > \frac{1}{2} - a$ prefer x to L_1. The set of voters who are indifferent between x and L_1 is given by voters satisfying $s^i \leq \frac{1}{2} - a$ and $(\frac{1}{2} + a - 2e^i)^{\frac{1}{2}} = (2s^i - \frac{1}{2} + a)^{\frac{1}{2}} + (2a)^{\frac{1}{2}}$. After some algebra, this condition can be rewritten as $s^i = \frac{1}{4} - \frac{a}{2} + \frac{1}{2}((\frac{1}{2} + a - 2e^i)^{\frac{1}{2}} - (2a)^{\frac{1}{2}})^2$. Since the utility of voting for x is increasing in s^i and constant in e^i, and the utility of voting for L_1 is decreasing in e^i and constant in s^i, voters with $s^i \leq \frac{1}{4} - \frac{a}{2} + \frac{1}{2}((\frac{1}{2} + a - 2e^i)^{\frac{1}{2}} - (2a)^{\frac{1}{2}})^2$ vote for L_1 and voters with $s^i > \frac{1}{4} - \frac{a}{2} + \frac{1}{2}((\frac{1}{2} + a - 2e^i)^{\frac{1}{2}} - (2a)^{\frac{1}{2}})^2$ vote for x.

Voters with $e^i > \frac{1}{4} + \frac{a}{2}$ and $s^i \leq \frac{1}{4} + \frac{a}{2}$: These voters have utility of voting for x given by

$$U(\frac{1}{2} - a, \frac{1}{2} - a, 0, \frac{1}{2} + a) = \begin{cases} (\frac{1}{2} - a)^{\frac{1}{2}} + (2a)^{\frac{1}{2}} & e^i > \frac{1}{2} - a \\ (2e^i - \frac{1}{2} + a)^{\frac{1}{2}} + (2a)^{\frac{1}{2}} & e^i \leq \frac{1}{2} - a \end{cases}$$

and utility of voting for y given by

$$U(\frac{1}{2} + a, \frac{1}{2} + a, 0, \frac{1}{2} + a) = \begin{cases} (\frac{1}{2} + a)^{\frac{1}{2}} & e^i > \frac{1}{2} + a \\ (2e^i - \frac{1}{2} - a)^{\frac{1}{2}} & e^i \leq \frac{1}{2} + a. \end{cases}$$

These voters also have utility of voting for L_1 given by

$$(\frac{1}{2} + a - 2s^i)^{\frac{1}{2}}.$$

It is immediately clear that voters with $e^i \leq \frac{1}{2} - a$ prefer voting for x to y. Voters with $e^i > \frac{1}{2} + a$ have utility $(\frac{1}{2} + a)^{\frac{1}{2}}$ of voting for y and $(\frac{1}{2} - a)^{\frac{1}{2}} + (2a)^{\frac{1}{2}}$ of voting for x. By concavity of $U(.)$, these voters prefer x to y as well. By continuity and monotonicity of $U(.)$, voters with $e^i \in (\frac{1}{2} - a, \frac{1}{2} + a]$ prefer x to y; thus, none of these voters vote for y.

Voters with $e^i > \frac{1}{2} - a$ have utility $(\frac{1}{2} - a)^{\frac{1}{2}} + (2a)^{\frac{1}{2}}$ of voting for x and $(\frac{1}{2} + a - 2s^i)^{\frac{1}{2}}$ of voting for L_1. Since $(\frac{1}{2} + a - 2s^i)^{\frac{1}{2}}$ is bounded above by $(\frac{1}{2} + a)^{\frac{1}{2}}$, these voters prefer x to L_1. Voters with $e^i \leq \frac{1}{2} - a$ have utility $(2e^i - \frac{1}{2} + a)^{\frac{1}{2}} + (2a)^{\frac{1}{2}}$ of voting for x and $(\frac{1}{2} + a - 2s^i)^{\frac{1}{2}}$ of voting for L_1. The set of voters who are indifferent between x and L_1 is therefore given by $s^i = \frac{1}{4} + \frac{a}{2} - \frac{1}{2}((2e^i - \frac{1}{2} + a)^{\frac{1}{2}} - (2a)^{\frac{1}{2}})^2$. Since the utility of voting for x is monotonically increasing in e^i and the utility of voting for L_1 is monotonically decreasing in s^i, voters with $s^i \leq \frac{1}{4} + \frac{a}{2} - \frac{1}{2}((2e^i - \frac{1}{2} + a)^{\frac{1}{2}} - (2a)^{\frac{1}{2}})^2$ vote for L_1 and voters with $s^i > \frac{1}{4} + \frac{a}{2} - \frac{1}{2}((2e^i - \frac{1}{2} + a)^{\frac{1}{2}} - (2a)^{\frac{1}{2}})^2$ vote for x.

Voters with $e^i > \frac{1}{4} + \frac{a}{2}$ and $s^i > \frac{1}{4} + \frac{a}{2}$: These voters have utility of voting for x given by

$$U(\frac{1}{2} - a, \frac{1}{2} - a, 0, 0) = \begin{cases} 2(\frac{1}{2} - a)^{\frac{1}{2}} & e^i > \frac{1}{2} - a,\ s^i > \frac{1}{2} - a \\ (\frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} + a)^{\frac{1}{2}} & e^i > \frac{1}{2} - a,\ s^i \leq \frac{1}{2} - a \\ (2e^i - \frac{1}{2} + a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}} & e^i \leq \frac{1}{2} - a,\ s^i > \frac{1}{2} - a \\ (2e^i - \frac{1}{2} + a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} + a)^{\frac{1}{2}} & e^i \leq \frac{1}{2} - a,\ s^i \leq \frac{1}{2} - a \end{cases}$$

Their utility of voting for y is given by

$$U(\frac{1}{2} + a, \frac{1}{2} + a, 0, 0) = \begin{cases} 2(\frac{1}{2} + a)^{\frac{1}{2}} & e^i > \frac{1}{2} + a,\ s^i > \frac{1}{2} + a \\ (\frac{1}{2} + a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} & e^i > \frac{1}{2} + a,\ s^i \leq \frac{1}{2} + a \\ (2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} + a)^{\frac{1}{2}} & e^i \leq \frac{1}{2} + a,\ s^i > \frac{1}{2} + a \\ (2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} & e^i \leq \frac{1}{2} + a,\ s^i \leq \frac{1}{2} + a \end{cases}$$

and their utility of voting for L_1 is $U(0, 0, 0, 0) = 0$.

First, note that none of these voters will vote for L_1. It is also clear that if $e^i > \frac{1}{2} + a$ and $s^i > \frac{1}{2} + a$, these voters prefer y to x. Similarly, if $e^i \leq \frac{1}{2} - a$, and $s^i \leq \frac{1}{2} - a$, these voters prefer x to y. Voters with $e^i > \frac{1}{2} + a$ and $\frac{1}{2} - a < s^i \leq \frac{1}{2} + a$ have utility $(\frac{1}{2} + a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y and $2(\frac{1}{2} - a)^{\frac{1}{2}}$ of voting for x and thus prefer voting for y. Symmetrically, voters with $s^i > \frac{1}{2} + a$ and $\frac{1}{2} - a < e^i \leq \frac{1}{2} + a$ prefer y as well. Voters with $e^i > \frac{1}{2} + a$ and $s^i \leq \frac{1}{2} - a$ have utility $(\frac{1}{2} + a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y and $(\frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} + a)^{\frac{1}{2}}$ of voting for x. These voters are indifferent between x and y if $a = \frac{1}{2}$; since $a < \frac{1}{6}$ by assumption, these voters prefer y to x. Symmetrically, voters with $s^i > \frac{1}{2} + a$ and $e^i \leq \frac{1}{2} - a$ prefer y to x as well.

Voters with $e^i \leq \frac{1}{2} - a$ and $\frac{1}{2} - a < s^i \leq \frac{1}{2} + a$ have utility $(2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y and $(2e^i - \frac{1}{2} + a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}}$ of voting for x. The set of indifferent voters is given by $s^i = \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((2s^i - \frac{1}{2} + a)^{\frac{1}{2}} - (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}})^2$. Since the utility of voting for y is monotonically decreasing in s^i and the utility of voting for x is increasing at a slower rate in e^i than the utility of voting for y, voters with $s^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((2s^i - \frac{1}{2} + a)^{\frac{1}{2}} - (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}})^2$ prefer voting for x and voters with $s^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((2s^i - \frac{1}{2} + a)^{\frac{1}{2}} - (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}})^2$ prefer voting for y.

Voters with $s^i \leq \frac{1}{2} - a$ and $\frac{1}{2} - a < e^i \leq \frac{1}{2} + a$ have utility $(2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y and $(\frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} + a)^{\frac{1}{2}}$ of voting for x. The set of indifferent voters is given by $e^i = \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((2s^i - \frac{1}{2} + a)^{\frac{1}{2}} - (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}})^2$. Since the utility of voting for y is monotonically increasing in e^i and monotonically increasing in s^i at a faster rate than the utility of voting for x, voters with $e^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((2s^i - \frac{1}{2} + a)^{\frac{1}{2}} - (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}})^2$ prefer voting for x and voters with $e^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}((2s^i - \frac{1}{2} + a)^{\frac{1}{2}} - (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}})^2$ prefer voting for y.

Voters with $\frac{1}{2} - a < s^i \leq \frac{1}{2} + a$ and $\frac{1}{2} - a < s^i \leq \frac{1}{2} + a$ have utility $(2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y and utility $2(\frac{1}{2} - a)^{\frac{1}{2}}$ of voting for x. The set of indifferent voters is given by $s^i = \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2$. Since the utility of voting for y is monotonically increasing in both s^i and e^i, and the utility of voting for x is constant in both variables, voters with $s^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2$ prefer voting for x. Voters with $s^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2$ prefer voting for y.

Consider now the case with $a > \frac{1}{6}$.

Voters with $e^i \leq \frac{1}{4} + \frac{a}{2}$ and $s^i \leq \frac{1}{4} + \frac{a}{2}$: These voters have utility of voting for x

given by

$$U(\frac{1}{2}-a,\frac{1}{2}-a,\frac{1}{2}+a,\frac{1}{2}+a)=\begin{cases}2(2a)^{\frac{1}{2}} & e^i \le \frac{1}{2}-a,\ s^i \le \frac{1}{2}-a\\ (2a)^{\frac{1}{2}}+(1-2s^i)^{\frac{1}{2}} & e^i \le \frac{1}{2}-a,\ s^i > \frac{1}{2}-a\\ (1-2e^i)^{\frac{1}{2}}+(2a)^{\frac{1}{2}} & e^i > \frac{1}{2}-a,\ s^i \le \frac{1}{2}-a\\ (1-2e^i)^{\frac{1}{2}}+(1-2s^i)^{\frac{1}{2}} & e^i > \frac{1}{2}-a,\ s^i > \frac{1}{2}-a\end{cases}$$

and their utility of voting for y is 0. None of these voters, therefore, vote for y. Their utility of voting for L_1 is given by

$$U(0,0,\frac{1}{2}-a,\frac{1}{2}-a)=(\frac{1}{2}+a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}+a-2s^i)^{\frac{1}{2}}$$

First, note that $(\frac{1}{2}+a-2x)^{\frac{1}{2}}$ is bounded above by $(1-2x)^{\frac{1}{2}}$. Thus, voters with $e^i > \frac{1}{2}-a$ and $s^i > \frac{1}{2}-a$ prefer x to L_1. Voters with $e^i \le \frac{1}{2}-a$ and $s^i \le \frac{1}{2}-a$ have utility $2(2a)^{\frac{1}{2}}$ of voting for x and utility $(\frac{1}{2}+a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}+a-2e^i)^{\frac{1}{2}}$ of voting for L_1. The set of voters who are indifferent between x and L_1 is given by $s^i=\frac{1}{4}+\frac{a}{2}-\frac{1}{2}(2(2a)^{\frac{1}{2}}-(\frac{1}{2}+a-2e^i)^{\frac{1}{2}})^2$. Since the utility of voting for L_1 is monotonically decreasing in both e^i and s^i and the utility of voting for x is constant, voters with $s^i \le \frac{1}{4}+\frac{a}{2}-\frac{1}{2}(2(2a)^{\frac{1}{2}}-(\frac{1}{2}+a-2e^i)^{\frac{1}{2}})^2$ vote for L_1 and voters with $s^i > \frac{1}{4}+\frac{a}{2}-\frac{1}{2}(2(2a)^{\frac{1}{2}}-(\frac{1}{2}+a-2e^i)^{\frac{1}{2}})^2$ vote for x.

Voters with $e^i \le \frac{1}{2}-a$ and $s^i > \frac{1}{2}-a$ have utility $(2a)^{\frac{1}{2}}+(1-2s^i)^{\frac{1}{2}}$ of voting for x and utility $(\frac{1}{2}+a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}+a-2s^i)^{\frac{1}{2}}$ of voting for L_1. The set of indifferent voters is given by $e^i=\frac{1}{4}+\frac{a}{2}-\frac{1}{2}((2a)^{\frac{1}{2}}+(1-2s^i)^{\frac{1}{2}}-(\frac{1}{2}+a-2s^i)^{\frac{1}{2}})^2$. Since the utility of voting for L_1 is monotonically decreasing in both e^i and s^i and the utility of voting for x is decreasing monotonically in s^i at a slower rate than L_1, voters with $e^i \le \frac{1}{4}+\frac{a}{2}-\frac{1}{2}((2a)^{\frac{1}{2}}+(1-2s^i)^{\frac{1}{2}}-(\frac{1}{2}+a-2s^i)^{\frac{1}{2}})^2$ vote for L_1. Voters with $e^i > \frac{1}{4}+\frac{a}{2}-\frac{1}{2}((2a)^{\frac{1}{2}}+(1-2s^i)^{\frac{1}{2}}-(\frac{1}{2}+a-2s^i)^{\frac{1}{2}})^2$ vote for x.

Voters with $s^i \le \frac{1}{2}-a$ and $e^i > \frac{1}{2}-a$ have utility $(2a)^{\frac{1}{2}}+(1-2e^i)^{\frac{1}{2}}$ of voting for x and utility $(\frac{1}{2}+a-2e^i)^{\frac{1}{2}}+(\frac{1}{2}+a-2e^i)^{\frac{1}{2}}$ of voting for L_1. The set of indifferent voters is given by $s^i = \frac{1}{4} + \frac{a}{2} - \frac{1}{2}((2a)^{\frac{1}{2}} + (1-2e^i)^{\frac{1}{2}} - (\frac{1}{2} + a - 2e^i)^{\frac{1}{2}})^2$. Since the utility of voting for L_1 is monotonically decreasing in both e^i and s^i and the utility of voting for x is decreasing monotonically in e^i at a slower rate than L_1, voters with $s^i \leq \frac{1}{4} + \frac{a}{2} - \frac{1}{2}((2a)^{\frac{1}{2}} + (1-2e^i)^{\frac{1}{2}} - (\frac{1}{2} + a - 2e^i)^{\frac{1}{2}})^2$ vote for L_1. Voters with $s^i > \frac{1}{4} + \frac{a}{2} - \frac{1}{2}((2a)^{\frac{1}{2}} + (1-2e^i)^{\frac{1}{2}} - (\frac{1}{2} + a - 2e^i)^{\frac{1}{2}})^2$ vote for x.

Voters with $e^i \leq \frac{1}{4} + \frac{a}{2}$ and $s^i > \frac{1}{4} + \frac{a}{2}$:

These voters have utility of voting for x given by

$$U(\frac{1}{2}-a, \frac{1}{2}-a, \frac{1}{2}+a, 0) = \begin{cases} (1-2e^i)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}} & e^i > \frac{1}{2}-a \\ (2a)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}-a \end{cases}$$

and utility of voting for y given by

$$U(\frac{1}{2}+a, \frac{1}{2}+a, \frac{1}{2}+a, 0) = \begin{cases} (\frac{1}{2}+a)^{\frac{1}{2}} & s^i > \frac{1}{2}+a \\ (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} & s^i \leq \frac{1}{2}+a \end{cases}$$

Finally, these voters have utility of voting for L_1 given by

$$U(0, 0, \frac{1}{2}+a, 0) = (\frac{1}{2}+a-2e^i)^{\frac{1}{2}}$$

First, note that x is preferred to L_1 by all voters. If $e^i > \frac{1}{2}-a$, the utility of voting for x is $(1-2e^i)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}}$ which is strictly larger than the utility of voting for L_1 for any $a < \frac{1}{2}$. If $e^i \leq \frac{1}{2}-a$, the utility of voting for x is $(2a)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}}$. If $e^i = 0$, the utility of voting for L_1 is $(\frac{1}{2}+a)^{\frac{1}{2}}$. By concavity of $U(.)$, x is preferred to L_1 for the voters with $e^i = 0$. Since the utility of voting for L_1 is monotonically decreasing in e^i and the utility of voting for x is constant in e^i, x is preferred to L_1 by all voters.

Note that the utility of voting for y is increasing in s^i for voters with $s^i \leq \frac{1}{2}+a$ and constant for voters with $s^i > \frac{1}{2}+a$. By the same argument as above, x is preferred to y when $e^i \leq \frac{1}{2}-a$ and $s^i > \frac{1}{2}+a$. Thus x is preferred to y for all voters with $e^i \leq \frac{1}{2}-a$.

Voters with $e^i > \frac{1}{2} - a$ have utility $(1-2e^i)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}}$ of voting for x. Consider voters with $s^i \leq \frac{1}{2}+a$. The set of indifferent voters is given by $s^i = \frac{1}{4}+\frac{a}{2}+\frac{1}{2}((1-2e^i)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}})^2$. Since the utility of voting for x is decreasing in e^i and the utility of voting for y is increasing in s^i, these voters vote for x if $s^i \leq \frac{1}{4}+\frac{a}{2}+\frac{1}{2}((1-2e^i)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}})^2$ and they vote for y if $s^i > \frac{1}{4}+\frac{a}{2}+\frac{1}{2}((1-2e^i)^{\frac{1}{2}}+(\frac{1}{2}-a)^{\frac{1}{2}})^2$.

Voters with $s^i > \frac{1}{2}+a$ have utility of $(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for y. The set of indifferent voters is given by $e^i = \frac{1}{2} - \frac{1}{2}((\frac{1}{2}+a)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$. Since the utility of voting for y is constant and the utility of voting for x is monotonically decreasing in e^i, voters with $e^i > \frac{1}{2} - \frac{1}{2}((\frac{1}{2}+a)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for y. Voters with $e^i \leq \frac{1}{2} - \frac{1}{2}((\frac{1}{2}+a)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for x.

Voters with $e^i > \frac{1}{4}+\frac{a}{2}$ and $s^i \leq \frac{1}{4}+\frac{a}{2}$:

These voters have utility of voting for x

$$U(\frac{1}{2}-a, \frac{1}{2}-a, 0, \frac{1}{2}+a) = \begin{cases} (\frac{1}{2}-a)^{\frac{1}{2}} + (1-2s^i)^{\frac{1}{2}} & s^i > \frac{1}{2}-a \\ (\frac{1}{2}-a)^{\frac{1}{2}} + (2a)^{\frac{1}{2}} & s^i \leq \frac{1}{2}-a \end{cases}$$

and utility of voting for y given by

$$U(\frac{1}{2}+a, \frac{1}{2}+a, 0, \frac{1}{2}+a) = \begin{cases} (\frac{1}{2}+a)^{\frac{1}{2}} & e^i > \frac{1}{2}+a \\ (2e^i - \frac{1}{2}-a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}+a \end{cases}$$

Their utility of voting for L_1 is given by

$$U(0,0,0,\frac{1}{2}+a) = (\frac{1}{2}+a-2s^i)^{\frac{1}{2}}$$

First note that x is preferred to L_1 by all of these voters for any $a < \frac{1}{2}$. If $s^i > \frac{1}{2}-a$, by inspection it is clear that x is preferred to L_1 for any $a < \frac{1}{2}$. The utility of voting for L_1 is bounded above by $(\frac{1}{2}+a)^{\frac{1}{2}}$. If $s^i \leq \frac{1}{2}-a$, then by concavity of $U(.)$ x is preferred to L_1.

For voters with $s^i \leq \frac{1}{2} - a$, x is preferred to y. The utility of voting for y is bounded above by $(\frac{1}{2}+a)^{\frac{1}{2}}$, and the utility of voting for x is $(\frac{1}{2}-a)^{\frac{1}{2}} + (2a)^{\frac{1}{2}}$. By concavity of $U(.)$, these voters prefer voting for x to y.

Voters with $s^i > \frac{1}{2} - a$ have utility $(\frac{1}{2}-a)^{\frac{1}{2}} + (1-2s^i)^{\frac{1}{2}}$ of voting for x. If $e^i > \frac{1}{2} + a$, these voters have utility $(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for y. Thus, the set of indifferent voters between x and y is given by $s^i = \frac{1}{2} - \frac{1}{2}((\frac{1}{2}+a)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$. Since the utility of voting for y is constant and the utility of voting for x is monotonically decreasing in s^i, voters with $s^i \leq \frac{1}{2} - \frac{1}{2}((\frac{1}{2}+a)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for x and voters with $s^i > \frac{1}{2} - \frac{1}{2}((\frac{1}{2}+a)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for y.

Voters with $s^i > \frac{1}{2} - a$ and $e^i \leq \frac{1}{2} + a$ have utility $(\frac{1}{2}-a)^{\frac{1}{2}} + (1-2s^i)^{\frac{1}{2}}$ of voting for x and $(2e^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y. The set of indifferent voters is given by $s^i = \frac{1}{2} - \frac{1}{2}((2e^i - \frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$. Since the utility of voting for y is monotonically increasing in e^i and the utility of voting for x is monotonically decreasing in s^i, voters with $s^i \leq \frac{1}{2} - \frac{1}{2}((2e^i - \frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for x. Voters with $s^i > \frac{1}{2} - \frac{1}{2}((2e^i - \frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2}-a)^{\frac{1}{2}})^2$ vote for y.

Voters with $e^i > \frac{1}{4} + \frac{a}{2}$ and $s^i > \frac{1}{4} + \frac{a}{2}$:

These voters have utility of voting for x given by

$$U(\frac{1}{2}-a, \frac{1}{2}-a, 0, 0) = 2(\frac{1}{2}-a)^{\frac{1}{2}}$$

and utility 0 of voting for L_1. Thus, none of these voters vote for L_1. Their utility of voting for y is given by

$$U(\frac{1}{2}+a, \frac{1}{2}+a, 0, 0) = \begin{cases} 2(\frac{1}{2}+a)^{\frac{1}{2}} & e^i > \frac{1}{2}+a,\ s^i > \frac{1}{2}+a \\ (\frac{1}{2}+a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} & e^i > \frac{1}{2}+a,\ s^i \leq \frac{1}{2}+a \\ (2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}+a,\ s^i > \frac{1}{2}+a \\ (2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + 2s^i - \frac{1}{2} - a)^{\frac{1}{2}} & e^i \leq \frac{1}{2}+a,\ s^i \leq \frac{1}{2}+a \end{cases}$$

Note that all voters with $e^i > \frac{1}{2} + a$ and $s^i > \frac{1}{2} + a$ vote for y. Voters with $e^i \leq \frac{1}{2} + a$ and $s^i > \frac{1}{2} + a$ have utility $(2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} + a)^{\frac{1}{2}}$ of voting for y and $2(\frac{1}{2} - a)^{\frac{1}{2}}$ of voting for x. The set of indifferent voters is given by $e^i = \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} + a)^{\frac{1}{2}})^2$. Since the utility of voting for x is constant and the utility of voting for y is increasing in e^i, these voters vote for x if $e^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} + a)^{\frac{1}{2}})^2$. They vote for y if $e^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} + a)^{\frac{1}{2}})^2$.

Voters with $e^i > \frac{1}{2} + a$ and $s^i \leq \frac{1}{2} + a$ have utility $2(\frac{1}{2} - a)^{\frac{1}{2}}$ of voting for x and utility $(\frac{1}{2} + a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y. The set of indifferent voters is given by $s^i = \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} + a)^{\frac{1}{2}})^2$. Since the utility of voting for x is constant and the utility of voting for y is increasing in s^i, these voters vote for x if $s^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} + a)^{\frac{1}{2}})^2$. They vote for y if $s^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (\frac{1}{2} + a)^{\frac{1}{2}})^2$.

Voters with $e^i \leq \frac{1}{2} + a$ and $s^i \leq \frac{1}{2} + a$ have utility $2(\frac{1}{2} - a)^{\frac{1}{2}}$ of voting for x and $(2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y. The set of indifferent voters is given by $s^i = \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2$. By the same argument as above, these voters vote for x if $s^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2$. They vote for y if $s^i > \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2} - a)^{\frac{1}{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2$. ∎

Lemma 8. Assume that there exist two candidate x and y with ideal policies $(\frac{1}{2} - a, \frac{1}{2} - a)$ and $(\frac{1}{2} + a, \frac{1}{2} + a)$, with $a \in [0, \frac{1}{2})$ respectively. Assume that there exist two candidates with ideal policies $(0, 0)$ and $(1, 1)$. Fix a citizen z with ideal policy (e^z, s^z) satisfying $e^z \in (0, \frac{1}{2} - a)$ and $s^z \in (0, \frac{1}{2} - a)$ or $e^z \in (\frac{1}{2} + a, 1)$ and $s^z \in (\frac{1}{2} + a, 1)$. For all citizens with ideal policy (e^i, s^i) satisfying $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$, and $\forall i$ such that $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$, i does not vote for z.

Proof: Let candidate z have ideal policy $(\frac{1}{2} - b, \frac{1}{2} - b)$ where $b \in (a, \frac{1}{2})$. The utility to voter i with $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$ of voting for candidate z is

$$U_i(e^z, s^z, \underline{e}^i, \underline{s}^i) = (\frac{3}{2} - b - 2e^i)^{\frac{1}{2}} + (\frac{1}{2} - b)^{\frac{1}{2}} \qquad e^i > \frac{1}{2} - b$$
$$U_i(e^z, s^z, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} + b)^{\frac{1}{2}} + (\frac{1}{2} - b)^{\frac{1}{2}} \qquad e^i \leq \frac{1}{2} - b$$

The same voters' utilities of voting for candidate x are

$$U_i(e^x, s^x, \underline{e}^i, \underline{s}^i) = (\frac{3}{2} - a - 2e^i)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}} \qquad e^i > \frac{1}{2} - a$$
$$U_i(e^x, s^x, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} + a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}} \qquad e^i \leq \frac{1}{2} - a$$

On inspection, it is clear that for all voters with $e^i > \frac{1}{2} - a$, $x > z$. Additionally, note that $(\frac{1}{2} + x)^{\frac{1}{2}} + (\frac{1}{2} - x)^{\frac{1}{2}}$ is monotonically decreasing in x. Thus, all voters with $e^i \leq \frac{1}{2} - b$ prefer x to z as well. It remains to be shown that voters with $e^i \in (\frac{1}{2} - b, \frac{1}{2} - a]$ prefer x to z. These voters have utility $(\frac{1}{2} + a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}}$ of voting for x, and utility $(\frac{3}{2} - b - 2e^i)^{\frac{1}{2}} + (\frac{1}{2} - b)^{\frac{1}{2}}$ of voting for z. Their utility of voting for z is bounded above by $(\frac{1}{2} + b)^{\frac{1}{2}} + (\frac{1}{2} - b)^{\frac{1}{2}}$. Since $(\frac{1}{2} + x)^{\frac{1}{2}} + (\frac{1}{2} - x)^{\frac{1}{2}}$ is monotonically decreasing in x and $b > a$ by assumption, these voters prefer x to z.

Now consider a voter i with $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$. The utility to voter i of voting for candidate x is

$$U_i(e^x, s^x, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} + a)^{\frac{1}{2}} \qquad \frac{1}{2} - a > s^i$$
$$U_i(e^x, s^x, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} - a)^{\frac{1}{2}} + (\frac{3}{2} - a - 2s^i)^{\frac{1}{2}} \qquad \frac{1}{2} - a \leq s^i$$

The utility to voter i of voting for candidate z is

$$U_i(e^z, s^z, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} - b)^{\frac{1}{2}} + (\frac{1}{2} + b)^{\frac{1}{2}} \qquad \frac{1}{2} - b > s^i$$
$$U_i(e^z, s^z, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} - b)^{\frac{1}{2}} + (\frac{3}{2} - b - 2s^i)^{\frac{1}{2}} \qquad \frac{1}{2} - b \leq s^i$$

As above, it is clear that for voters with $s^i > s^x$ or $s^i \leq s^z$, candidate x is preferred to candidate z. Voters with $s^i \in (\frac{1}{2} - b, \frac{1}{2} - a)$ obtain utility $(\frac{1}{2} - b)^{\frac{1}{2}} + (\frac{3}{2} - b - 2s^i)^{\frac{1}{2}}$ of voting for candidate z and utility $(\frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} + a)^{\frac{1}{2}}$ of voting for candidate x. Note that $(\frac{1}{2} - b)^{\frac{1}{2}} + (\frac{3}{2} - b - 2s^i)^{\frac{1}{2}}$ is bounded above by $(\frac{1}{2} - b)^{\frac{1}{2}} + (\frac{1}{2} + b)^{\frac{1}{2}}$. Since $b < a$ by assumption, and $(\frac{1}{2} + x)^{\frac{1}{2}} + (\frac{1}{2} - x)^{\frac{1}{2}}$ is monotonically decreasing in x, these voters prefer x to z as well.

To complete the proof of the lemma, it remains to be shown that candidate y, having ideal policy $(\frac{1}{2}+a, \frac{1}{2}+a)$ is strictly preferred by these voters to any candidate z with ideal policy $(\frac{1}{2}+b, \frac{1}{2}+b)$, where $b > a$. The utility to voter i with $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$ of voting for candidate y is

$$U_i(e^y, s^y, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}} \qquad s^i \leq \frac{1}{2} + a$$
$$U_i(e^y, s^y, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} + a)^{\frac{1}{2}} \qquad s^i > \frac{1}{2} + a$$

The same voter's utility of voting for candidate z is

$$U_i(e^z, s^z, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} - b)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - b)^{\frac{1}{2}} \qquad s^i \leq \frac{1}{2} + b$$
$$U_i(e^z, s^z, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} - b)^{\frac{1}{2}} + (\frac{1}{2} + b)^{\frac{1}{2}} \qquad s^i > \frac{1}{2} + b$$

For all i with $s^i \leq \frac{1}{2}+a$ and $s^i > \frac{1}{2}+b$, it is clear that voting for y is preferred to voting for z. For voters with $s^i \in (\frac{1}{2}+a, \frac{1}{2}+b]$, voting for y is preferred if $(\frac{1}{2}-a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}} \geq (\frac{1}{2}-b)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - b)^{\frac{1}{2}}$. Note that the right hand side of this inequality is bounded above by $(\frac{1}{2}-b)^{\frac{1}{2}} + (\frac{1}{2}+b)^{\frac{1}{2}}$. Thus, all voters with ideal policies satisfying $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$ prefer to vote for y instead of z.

Consider now voters with ideal policy $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$. These voters' utility of voting for candidate z is

$$U_i(e^z, s^z, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} + b)^{\frac{1}{2}} + (\frac{1}{2} - b)^{\frac{1}{2}} \qquad e^i > \frac{1}{2} + b$$
$$U_i(e^z, s^z, \underline{e}^i, \underline{s}^i) = (2e^i - \frac{1}{2} - b)^{\frac{1}{2}} + (\frac{1}{2} - b)^{\frac{1}{2}} \qquad e^i \leq \frac{1}{2} + b$$

These voters utility of voting for candidate y is

$$U_i(e^y, s^y, \underline{e}^i, \underline{s}^i) = (\frac{1}{2} + a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}} \qquad e^i > \frac{1}{2} + a$$
$$U_i(e^y, s^y, \underline{e}^i, \underline{s}^i) = (2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2} - a)^{\frac{1}{2}} \qquad e^i \leq \frac{1}{2} + a$$

Clearly, for any voters with $e^i > \frac{1}{2}+b$ or $e^i \leq \frac{1}{2}+a$, voting for y is preferred to voting for z. Voters with $e^i \in (\frac{1}{2}+a, \frac{1}{2}+b]$ obtain utility of $(\frac{1}{2}+a)^{\frac{1}{2}} + (\frac{1}{2}-a)^{\frac{1}{2}}$ from voting for y, and utility $(2e^i - \frac{1}{2} - b)^{\frac{1}{2}} + (\frac{1}{2} - b)^{\frac{1}{2}}$ from voting for z. Since $(2e^i - \frac{1}{2} - b)^{\frac{1}{2}} + (\frac{1}{2} - b)^{\frac{1}{2}}$ is bounded above by $(\frac{1}{2} + b)^{\frac{1}{2}} + (\frac{1}{2} - b)^{\frac{1}{2}}$, these voters prefer voting for y as well. ∎

Lemma 9. Assume there exist two candidates x and y with ideal policies $(\frac{1}{2}-a, \frac{1}{2}-a)$ and $(\frac{1}{2}+a, \frac{1}{2}+a)$ respectively. Assume that there exist two candidates with ideal policies $(0,0)$ and $(1,1)$. Let $\bar{s} = \bar{e} = \frac{1}{2}(\frac{2}{\sqrt{2}} - (\frac{1}{2}+a)^{\frac{1}{2}})^2 + \frac{1}{4} + \frac{a}{2}$. An entrant with ideal policy $(\frac{1}{2}, \frac{1}{2})$ obtains all votes from voters with ideal policies satisfying $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$, and from all voters with ideal policies satisfying $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$. An entrant with ideal policy $(\frac{1}{2}, \frac{1}{2})$ also obtains votes from voters with ideal policies satisfying $e^i \in [1-\bar{e}, \bar{e}]$ or $s^i \in [1-\bar{s}, \bar{s}]$. Finally, voters with ideal policies satisfying $e^i \in [\frac{1}{2}-a, 1-\bar{e}]$ and $s^i \notin [1-\bar{s}, \bar{s}]$ vote for the entrant if $s^i \geq \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(\frac{2}{\sqrt{2}} - (\frac{3}{2} - 2e^i - a)^{\frac{1}{2}})^2$ and voters with ideal policies satisfying $e^i \in [\bar{e}, \frac{1}{2}+a]$ and $s^i \notin [1-\bar{s}, \bar{s}]$ vote for the entrant if $s^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(\frac{2}{\sqrt{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2$.

Proof: First, note that all voters have utility $\frac{2}{\sqrt{2}}$ of voting for a candidate with ideal policy $(\frac{1}{2}, \frac{1}{2})$. By Lemma 8, the median entrant obtains all votes from citizens with ideal policies (e^i, s^i) satisfying $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ or $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$.

Consider citizens with ideal policies satisfying $e^i \geq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$. Note that all of these citizens prefer y to x. These citizens can be broken up into four subsets by whether $e^i \geq \frac{1}{2}+a$, $s^i \geq \frac{1}{2}+a$, both, or neither. Citizens with both e^i and $s^i \geq \frac{1}{2}+a$ have utility of $2(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for y, and therefore the median entrant receives no votes from this group. Consider now the subset of citizens with $e^i \geq \frac{1}{2}+a$ and $s^i \leq \frac{1}{2}+a$. These citizens have utility $(\frac{1}{2}+a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y and $\frac{2}{\sqrt{2}}$ of voting for the median entrant. The set of indifferent citizens are those with

$s^i = \bar{s} = \frac{1}{2}(\frac{2}{\sqrt{2}} - (\frac{1}{2}+a)^{\frac{1}{2}})^2 + \frac{1}{4} + \frac{a}{2}$. Since the utility of voting for y is increasing in s^i, all citizens with $e^i \geq \frac{1}{2}+a$ and $s^i \leq \bar{s}$ vote for the median entrant. Similarly, the subset of citizens with $e^i \leq \frac{1}{2}+a$ and $s^i \geq \frac{1}{2}+a$ have utility $(2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for y and $\frac{2}{\sqrt{2}}$ of voting for the median entrant. The set of indifferent citizens are those with $e^i = \bar{e} = \frac{1}{2}(\frac{2}{\sqrt{2}} - (\frac{1}{2}+a)^{\frac{1}{2}})^2 + \frac{1}{4} + \frac{a}{2}$. Since the utility of voting for y is increasing in e^i, all citizens with $s^i \geq \frac{1}{2}+a$ and $e^i \leq \bar{e}$ vote for the median entrant. Finally, citizens with $s^i \leq \frac{1}{2}+a$ and $e^i \leq \frac{1}{2}+a$ have utility $(2e^i - \frac{1}{2} - a)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - a)^{\frac{1}{2}}$ of voting for y and $\frac{2}{\sqrt{2}}$ of voting for the median entrant. Thus, the set of indifferent citizens are given by $s^i(e^i) = \frac{1}{2}(\frac{2}{\sqrt{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2 + \frac{1}{4} + \frac{a}{2}$. Since the utility of voting for y is increasing in e^i and s^i, all citizens satisfying $s^i \leq \frac{1}{2}+a$ and $e^i \leq \frac{1}{2}+a$ with $s^i \leq \frac{1}{2}(\frac{2}{\sqrt{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2 + \frac{1}{4} + \frac{a}{2}$ vote for z, while those with $s^i > \frac{1}{2}(\frac{2}{\sqrt{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2 + \frac{1}{4} + \frac{a}{2}$ vote for y.

Now consider citizens with ideal policies satisfying $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$. These citizens all prefer x to y. Similarly, they can be broken up into four subsets by whether $e^i \leq \frac{1}{2} - a$, $s^i \leq \frac{1}{2} - a$, both or neither. Citizens with both e^i and $s^i \leq \frac{1}{2} - a$ have utility $2(\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for x, and therefore prefer x to z. Citizens with $e^i \leq \frac{1}{2} - a$ and $s^i > \frac{1}{2} - a$ have utility $(\frac{3}{2} - 2s^i - a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for x. The set of indifferent citizens are those with $s^i = 1 - \bar{s} = \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(\frac{2}{\sqrt{2}} - (\frac{1}{2}+a)^{\frac{1}{2}})^2$. Since the utility of voting for x is decreasing in s^i, all citizens with $e^i \leq \frac{1}{2} - a$ and $s^i \geq 1 - \bar{s}$ vote for the entrant, while those with $e^i \leq \frac{1}{2} - a$ and $s^i < 1 - \bar{s}$ vote for x. Citizens with $s^i \leq \frac{1}{2} - a$ and $e^i > \frac{1}{2} - a$ have utility $(\frac{3}{2} - 2e^i - a)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}}$ of voting for x. The set of indifferent citizens is given by $e^i = 1 - \bar{e} = \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(\frac{2}{\sqrt{2}} - (\frac{1}{2}+a)^{\frac{1}{2}})^2$. Since the utility of voting for x is decreasing in e^i, all citizens with $s^i \leq \frac{1}{2} - a$ and $e^i \geq 1 - \bar{e}$ vote for z, whereas citizens with $s^i \leq \frac{1}{2} - a$ and $e^i < 1 - \bar{e}$ vote for x. Finally, citizens with $s^i > \frac{1}{2} - a$ and $e^i > \frac{1}{2} - a$ obtain utility $(\frac{3}{2} - 2s^i - a)^{\frac{1}{2}} + (\frac{3}{2} - 2e^i - a)^{\frac{1}{2}}$ of voting for x. The set of indifferent citizens is given by $s^i(e^i) = \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(\frac{2}{\sqrt{2}} - (\frac{3}{2} - 2e^i - a)^{\frac{1}{2}})^2$. All citizens with s^i and e^i satisfying $s^i \geq \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(\frac{2}{\sqrt{2}} - (\frac{3}{2} - 2e^i - a)^{\frac{1}{2}})^2$ vote for z, whereas those with

e^i and s^i satisfying $s^i < \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(\frac{2}{\sqrt{2}} - (\frac{3}{2} - 2e^i - a)^{\frac{1}{2}})^2$ vote for x. ■

Lemma 10. Assume there exist two candidates x and y with ideal policies $(\frac{1}{2}-a, \frac{1}{2}-a)$ and $(\frac{1}{2}+a, \frac{1}{2}+a)$ respectively. Assume that there exist two candidates with ideal policies $(0,0)$ and $(1,1)$. An entrant z with ideal policy $(\frac{1}{2}-b, \frac{1}{2}-b)$, where $b \in (0,a)$, obtains votes from all voters with (e^i, s^i) satisfying $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ or $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$. Let $\underline{e} = \underline{s} = \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(2(\frac{1}{2}+b)^{\frac{1}{2}} - (\frac{1}{2}+a)^{\frac{1}{2}})^2$. Let $\bar{e} = \bar{s} = \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2}-b)^{\frac{1}{2}} - (\frac{1}{2}+a)^{\frac{1}{2}})^2$. Additionally, all voters with ideal policies satisfying $e^i \in [\underline{e}, \bar{e}]$ or $s^i \in [\underline{s}, \bar{s}]$ vote for z. Finally, voters with $e^i \in [\frac{1}{2}-a, \underline{e}]$ vote for z if $s^i \geq \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(2(\frac{1}{2}+b)^{\frac{1}{2}} - (\frac{3}{2} - 2e^i - a)^{\frac{1}{2}})^2$, while voters with $e^i \in [\bar{e}, \frac{1}{2}+a]$ vote for z if $s^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2}-b)^{\frac{1}{2}} - (2e^i - \frac{1}{2} - a)^{\frac{1}{2}})^2$.

Proof: Follows the exact same argument as Lemma 9. The only difference is the utility voters have for z. Voters with ideal policies satisfying $e^i \geq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$ have utility $2(\frac{1}{2}-b)^{\frac{1}{2}}$ of voting for z, whereas voters with ideal policies satisfying $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ have utility of voting for z given by

$$U(e^z, s^z, \underline{e}^i, \underline{s}^i) = \begin{cases} (\frac{3}{2} - 2e^i - b)^{\frac{1}{2}} + (\frac{3}{2} - 2s^i - b)^{\frac{1}{2}} & e^i \geq \frac{1}{2} - b,\ s^i \geq \frac{1}{2} - b \\ (\frac{3}{2} - 2e^i - b)^{\frac{1}{2}} + (\frac{1}{2} + b)^{\frac{1}{2}} & e^i \geq \frac{1}{2} - b,\ s^i < \frac{1}{2} - b \\ (\frac{1}{2} + b)^{\frac{1}{2}} + (\frac{3}{2} - 2s^i - b)^{\frac{1}{2}} & e^i < \frac{1}{2} - b,\ s^i \geq \frac{1}{2} - b \\ 2(\frac{1}{2} + b)^{\frac{1}{2}} & e^i < \frac{1}{2} - b,\ s^i < \frac{1}{2} - b \end{cases}$$

As before, by Lemma 8 voters with $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ and voters with $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$ vote for z. ■

Lemma 11. Assume there exist two candidates x and y with ideal policies $(\frac{1}{2}-a, \frac{1}{2}-a)$ and $(\frac{1}{2}+a, \frac{1}{2}+a)$ respectively. Assume that there exist two candidates with ideal policies $(0,0)$ and $(1,1)$. An entrant z with ideal policy $(\frac{1}{2}+b, \frac{1}{2}+b)$, where $b \in (0,a)$ obtains all votes from voters with $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ and from voters with $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$. Let $\underline{e} = \underline{s} = \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(2(\frac{1}{2}-b)^{\frac{1}{2}} + (\frac{1}{2}+a)^{\frac{1}{2}})^2$. Let $\bar{s} = \bar{e} = \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2}+b)^{\frac{1}{2}} - (\frac{1}{2}+a)^{\frac{1}{2}})^2$. Additionally, voters with $e^i \in [\underline{e}, \bar{e}]$ or $s^i \in [\underline{s}, \bar{s}]$ vote for z. Finally, voters with $e^i \in [\frac{1}{2}-a, \underline{e}]$ vote for z if $s^i \geq \frac{3}{4} - \frac{a}{2} - \frac{1}{2}(2(\frac{1}{2}-b)^{\frac{1}{2}} - (\frac{3}{2} - 2e^i - a)^{\frac{1}{2}})^2$, whereas those with $s^i \in [\bar{s}, \frac{1}{2}+a]$ vote for z if $s^i \leq \frac{1}{4} + \frac{a}{2} + \frac{1}{2}(2(\frac{1}{2}+b)^{\frac{1}{2}} - (\frac{1}{2}+a)^{\frac{1}{2}})^2$.

Proof: Follows the same argument as Lemma 9. The only difference is the utility voters have for z. Voters with ideal policies satisfying $e^i \leq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ have utility $2(\frac{1}{2} - b)^{\frac{1}{2}}$ of voting for z, whereas voters with ideal policies satisfying $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ have utility of voting for z given by

$$U(e^z, s^z, \underline{e}^i, \underline{s}^i) = \begin{cases} 2(\frac{1}{2} + b)^{\frac{1}{2}} & e^i \geq \frac{1}{2} + b,\ s^i \geq \frac{1}{2} + b \\ (\frac{1}{2} + b)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - b)^{\frac{1}{2}} & e^i \geq \frac{1}{2} + b,\ s^i < \frac{1}{2} + b \\ (2e^i - \frac{1}{2} - b)^{\frac{1}{2}} + (\frac{1}{2} + b)^{\frac{1}{2}} & e^i < \frac{1}{2} + b,\ s^i \geq \frac{1}{2} + b \\ (2e^i - \frac{1}{2} - b)^{\frac{1}{2}} + (2s^i - \frac{1}{2} - b)^{\frac{1}{2}} & e^i < \frac{1}{2} + b,\ s^i < \frac{1}{2} + b \end{cases}$$

As before, by Lemma 8 voters with $e^i \geq \frac{1}{2}$ and $s^i \leq \frac{1}{2}$ and voters with $e^i \leq \frac{1}{2}$ and $s^i \geq \frac{1}{2}$ vote for z. ■

Proof of Proposition 4: Lemmas 3 through 11 characterize vote shares for candidates in equilibrium, for any linear entrant, and if any equilibrium candidate exits. Assume $F(e^i, s^i)$ is symmetric over the line $s^i = 1 - e^i$.

Suppose there exists four candidates, L_1, L_2, x, and y with ideal policies $(0,0)$, $(1,1)$, $(\frac{1}{2} - a, \frac{1}{2} - a)$, and $(\frac{1}{2} + a, \frac{1}{2} + a)$ respectively. In order for this to be an equilibrium where candidates L_1 and L_2 lose with certainty, candidates x and y must both win with positive probability. If not, by exiting, they would guarantee the victory of the other moderate, and thus have no effect on the outcome of the election, but save cost c. In equilibrium, candidates x and y receive a payoff of $\frac{1}{2}b - c - \frac{1}{2}d(x,y)$. By exiting, each candidate would receive $-d(x,y)$. Therefore, in order to sustain this equilibrium, we require $b \geq 2c - d(x,y)$.

Candidates L_1 and L_2 obtain a payoff of $-\frac{1}{2}(d(L_i, x) + d(L_i, y)) - c$ in equilibrium. Since the equilibrium is fully symmetric, consider, without loss of generality, candidate L_1. By exiting, L_1 shifts the frame of reference for all voters who have $e^i \geq \frac{1}{2}$ or $s^i \geq \frac{1}{2}$. As noted above in Lemma **??**, by exiting candidate L_1 tilts the perception of y favorably relative to x for all voters whose frame changes. Candidate x, however, obtains votes from all voters who previously voted for L_1. If the set of voters who would have voted for L_1 is larger than the set of voters for whom the change in frame causes them to vote for y, then L_1 obtains $-d(L_1, x)$ by exiting, which is strictly larger and thus this cannot be an equilibrium. Therefore, only distributions and candidate configurations which satisfy $v_y(L_2, x, y) > v_x(L_2, x, y)$ and $v_x(L_1, x, y) > v_y(L_1, x, y)$ can support this equilibrium. If those conditions are satisfied, L_1 obtains $-d(L_1, y)$ by exiting. Since this is a linear equilibrium, $d(L_1, y)$ can be decomposed as $d(L_1, x) + d(x, y)$. Thus, in order for L_1 and L_2 to find it optimal to enter, it must be the case that $-\frac{1}{2}[2d(L_1, x) + d(x, y)] - c \geq -d(L_1, x) - d(x, y)$, or $\frac{3}{2}d(x, y) \geq c$.

Finally, if any potential candidate can enter and win the election, they obtain a payoff of $b - c$ by doing so. In equilibrium, they would obtain a payoff of $-\frac{1}{2}[d(x, i) + d(y, i)]$. They will not find it optimal to enter if $b \geq c - \frac{1}{2}d(x, i) - \frac{1}{2}d(y, i)$. No other sure losers would like to enter. The frames are invariant to any other entrant since the distribution is bounded at 0 and 1 in each dimension. If they are extreme with respect to party competition, they exclusively cannibalize votes from their preferred moderate. Since $F(.)$ is symmetric over the set of voters indifferent between x and y, any centrist sure loser who has preferences over x and y extracts strictly more votes from her preferred moderate. ∎

Appendix C

Appendices to Giving the Gift of Guilt Avoidance

C.1 Proof of Representation Theorem

C.1.1 Preliminaries

A set of linear utility functions $\mathcal{F}$ is redundant if one of them is constant, or if there exists $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $f = \alpha g + \beta$ for some $f, g \in \mathcal{F}$. By Herstein and Milnor (1953), the latter condition holds if and only if f and g represent the same preference ranking.

Lemma 1. If $S \geq 2$ and functions $u_1, \ldots, u_S \in \mathcal{U}$ are not redundant, then there are $x_1^*, \ldots, x_S^* \in X$ such that $u_i(x_i^*) > u_i(x_j^*)$ for all $i \neq j$.

Proof. This proof is due to Kopylov (2009a). Take any $k, \ell \in S$ such that $k > \ell$. Then u_k and u_ℓ are non-constant and represent different rankings of X. Without loss of generality, $u_k(x) \geq u_k(y)$ and $u_\ell(y) > u_{|ell}(x)$ for some $x, y \in X$. Take $x', y' \in X$ such that $u_k(x') > u_k(y')$. Take $\varepsilon > 0$ such that $u_\ell(\varepsilon y' + (1-\varepsilon)y) > u_\ell(\varepsilon x' + (1-\varepsilon)x)$. Let $x_{k\ell} = \varepsilon x' + (1-\varepsilon)x$ and $x_{\ell k} = \varepsilon y' + (1-\varepsilon)y$. Then $u_k(x_{k\ell}) > u_k(x_{\ell k})$ and $u_\ell(x_{\ell k}) > u_\ell(x_{k\ell})$. For any $i \in S$, let

$$x_i^* = \sum_{k,\ell \in S, k>\ell} \frac{2}{S(S-1)} x_{k\ell}^i,$$

where $x_{k\ell}^i = x_{k\ell}$ if $u_i(x_{k\ell}) \geq u_i(x_{\ell k})$ and $x_{k\ell}^i = x_{\ell k}$ otherwise. Then for any $i, j \in S$ such that $i \neq j$,

$$u_i(x_i^*) = \sum_{k,\ell \in S, k>\ell} \frac{2}{S(S-1)} u_i\left(x_{k\ell}^i\right) > \sum_{k,\ell \in S, k>\ell} \frac{2}{S(S-1)} u_i\left(x_{k\ell}^j\right) = u_i\left(x_j^*\right)$$

because $u_i(x_{ij}) > u_i(x_{ji})$ and $u_i(x_{k\ell}^i) \geq u_i(x_{k\ell}^j)$ for any $k, \ell \in S$ such that $k > \ell$. □

C.1.2 Axioms Imply Representation

Lemma 2. The metric space $(\mathcal{M}_1, \mu_1)$, where μ_1 is the Hausdorff metric defined in section 1.3.1, satisfies the following properties:

1. $\mathcal{M}_1$ is compact

2. mixture operation mapping $[0,1] \times \mathcal{M}_1 \times \mathcal{M}_1 \to X$ is continuous

3. $\forall\ \alpha \in [0,1]$ and $a, b, a', b' \in \mathcal{M}_1$

$$\mu_1(\alpha a + (1-\alpha)a, \alpha b + (1-\alpha)b) \le \max\{\mu_1(a,b), \mu_1(a',b')\}$$

Proof. Take $a, b, a', b' \in \mathcal{M}_1$. Suppose $\max\{\mu_1(a,b), \mu_1(a',b')\} = \varepsilon$. Then $b \subset N_\varepsilon(a)$, and $a \subset N_\varepsilon(b)$, where

$$\begin{aligned} N_\varepsilon(a) &= \{x \in X | d(x,a) \le \varepsilon\} \\ d(x,a) &= \min_{y \in a} d(x,y) \end{aligned}$$

Given $y \in b$, there exists $x \in a$ such that $d(x,y) \le \varepsilon$.
Given $y' \in b'$, there exists $x' \in a'$ such that $d(x',y') \le \varepsilon$.
Therefore, $d(\alpha x + (1-\alpha)x', \alpha y + (1-\alpha)y') \le \varepsilon$, which implies

$$\{\alpha y + (1-\alpha)y' | y \in b,\ y' \in b'\} \subset N_\varepsilon\{\alpha x + (1-\alpha)x' | x \in a,\ x' \in a'\}$$

And the same is true in reverse, thus $\mu_1(\alpha a + (1-\alpha)a', \alpha b + (1-\alpha)b') \le \varepsilon$. □

By Lemma 2, *Partial Set Betweenness*, and Corollary 3.3 of Theorem 2.1 in Kopylov (2009a), this has utility representation of the form

$$W(A) = \max_{a \in A} w(a) - \max_{b \in A} v_1(b) - \max_{c \in A} v_2(c),$$

for some $w, v_1, v_2 \in \mathcal{U}$, where $\mathcal{U}$ is the set of all continuous linear functions $u : X \to \mathbb{R}$.

Lemma 3. Given *Partial Set-Betweenness*, the preceding representation can be written in the form

$$W(A) = \max_{a \in A} [(1+\kappa) U(a) + V(a)] - \max_{b \in a} V(b) - \kappa \max_{c \in A} U(c)$$

for some $v \in \mathcal{U}$ and $\kappa \geq 0$.

Proof. Let $U = w - v_1 - v_2$. Suppose first that v_1, v_2, U are redundant. This implies one of the following cases:

1. v_1 and v_2 are redundant. Then for each $A \in \mathcal{M}_0$,

   $$W(A) = \max_{a \in A} w(a) - \max_{b \in A} [v_1(b) + v_2(b)]$$

   This is equivalent to the desired form for $V = v_1 + v_2$ and $\kappa = 0$.

2. $v_1 = \alpha U + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$. Then the utility function has the desired form for $V = v_2$ and $\kappa = \alpha$:

   $$W(A) = \max_{a \in A} [U(a) + \kappa U(a) + \beta + V(a)] - \max_{c \in A} [\kappa U(b) - \beta] - \max_{b \in A} V(b)$$

3. $v_2 = \alpha U + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$. This is equivalent to the preceding case; the utility function has the desired form for $v = v_1$ and $\kappa = \alpha$.

4. v_1 and v_2 are not redundant, and U is constant. In this case the preferences will violate *Partial Set-Betweenness.* To see, take $a, b \in \mathcal{M}_1$ s.t. $v_1(a) > v_1(b)$ and $v_2(b) > v_2(a)$. Without loss of generality $w(a) \geq w(b)$. Let $c = \frac{1}{2}a + \frac{1}{2}b$. Then $v_1(a) > v_1(c) > v_1(b)$ and $v_2(b) > v_2(c) > v_2(a)$. Thus

   $$\begin{aligned} W(\{a,c\}) &= w(a) - v_1(a) - v_2(c) > w(a) - v_1(a) - v_2(b) = W(\{a,b,c\}) \\ W(\{y,z\}) &= w(c) - v_1(c) - v_2(b) > w(a) - v_1(a) - v_2(b) = W(\{a,b,c\}) \end{aligned}$$

   because $w(c) - v_1(c) = v_2(c) + u(c) > v_2(a) + u(a) = w(a) - v_1(a)$. Rankings $\{a,c\} \succ \{a,b,c\}$ and $\{b,c\} \succ \{a,b,c\}$ violate *Partial Set-Betweenness* because $\{a,c\}, \{b,c\} \in \mathcal{M}_c$.

Suppose U, v_1, v_2 are not redundant. Then one of the following must hold.

1. w, v_1, v_2, U are all not redundant. This will violate *Partial Set-Betweenness*. By Lemma 1, there are $a, b_1, b_2, c \in \mathcal{M}_1$ such that

$$\begin{aligned} w(a) &> \max\{w(b_1),\ w(b_2),\ w(c)\} \\ v_1(b_1) &> \max\{v_1(a),\ v_1(b_2),\ v_1(c)\} \\ v_2(b_2) &> \max\{v_2(a),\ v_2(b_1),\ v_2(c)\} \\ U(c) &> \max\{U(a),\ U(b_1),\ U(b_2)\} \end{aligned}$$

   Let $A = \{a, b_1, c\}$ and $B = \{a, b_2, c\}$. Then both utilities

$$\begin{aligned} W(A) &= w(a) - v_1(b_1) - \max\{v_2(a), v_2(b_1), v_2(c)\} \text{ and} \\ W(B) &= w(a) - \max\{v_1(a), v_1(b_2), v_1(c)\} - v_2(b_2) \end{aligned}$$

   are strictly greater than the utility

$$W(A \cup B) = w(a) - v_1(b_1) - v_2(b_2)$$

   Yet the rankings $A \succ A \cup B$ and $B \succ A \cup B$ violate *Partial Set-Betweenness* because $A, B \in \mathcal{M}_c$.

2. $w = \alpha U + \beta$ for some $\alpha \geq 0$ and $\beta \in \mathbb{R}$. This, too, will violate *Partial Set-Betweenness*. By Lemma 1, there are $b_1, b_2, c \in \mathcal{M}_1$ s.t.

$$\begin{aligned} v_1(b_1) &> \max\{v_1(b_2),\ v_1(c)\} \\ v_2(b_2) &> \max\{v_2(b_1),\ v_2(c)\} \\ u(c) &> \max\{u(b_1),\ u(b_2)\} \end{aligned}$$

   Then $w(c) > \max\{w(b_1), w(b_2)\}$. Let $A = \{b_1, c\}$ and $B = \{b_2, c\}$. Then

$$\begin{aligned} W(A) &= w(c) - v_1(b_1) - \max\{v_2(b_1), v_2(c)\} > w(c) - v_1(b_1) - v_2(b_2) = W(A \cup B) \\ W(B) &= w(c) - \max\{v_1(b_2), v_1(c)\} - v_2(b_2) > w(c) - v_1(b_1) - v_2(b_2) = W(A \cup B) \end{aligned}$$

But, again, $A \succ A \cup B$ and $B \succ A \cup B$ violate *Partial Set-Betweenness* because $A, B \in \mathcal{M}_c$.

3. $w = \alpha v_1 + \beta$ for some $\alpha \geq 0$ and $\beta \in \mathbb{R}$. If $\alpha \geq 1$, then $\forall\ A \in \mathcal{M}_0$,

$$W(A) = \max_{a \in A} [w(a) - v_1(a)] - \max_{b \in A} v_2(b)$$

has the required form with $V = v_2$ and $\kappa = 0$.

$$\begin{aligned} W(A) &= \max_{a \in A} [U(a) + v_1(a) + V(a) - v_1(a)] - \max_{b \in A} V(b) \\ &= \max_{a \in A} [U(a) + V(a)] - \max_{b \in A} V(b) \end{aligned}$$

If $\alpha < 1$, then $\forall\ A \in \mathcal{M}_0$,

$$W(A) = \max_{a \in A} w'(a) - \max_{b \in A} v_1'(b) - \max_{c \in A} v_2(c)$$

where $w' = 0$ and $v_1' = v_1 - w = (1-\alpha)v_1 - \beta$.[1] Then $w' = 0 \cdot u$, which implies v_1', v_2, u are not redundant. By case two, this contradicts *Partial Set-Betweenness*.

4. $w = \alpha v_2 + \beta$ for some $\alpha \geq 0$ and $\beta \in \mathbb{R}$. This case is analogous to the previous case.

Ergo, representation exists. □

For the singleton menu of menus $A = a$, $W(A) = U(a)$. As singleton menus of menus satisfy *Set-Betweenness*, by Gul and Pesendorfer (2001),

$$U(a) = \max_{x \in a} [u(x) + v(x)] - \max_{y \in a} v(y),$$

where u and v are linear, positive functions.

[1] If $\alpha > 1$, this is not a positive function.

C.1.3 Representation Implies Axioms

The usual argument implies *Weak Order* and *Continuity*.

To show *Independence*, recall $\alpha A + (1-\alpha)C = \{\alpha a + (1-\alpha)c | a \in A, c \in C\}$

$$\begin{aligned} W(\alpha A + (1-\alpha)C) &= \max_{a \in A, c \in C} \left[(1+\kappa)U(\alpha a + (1-\alpha)c) + V(\alpha a + (1-\alpha)c)\right] \\ &\quad - \max_{b \in A, d \in C} V(\alpha b + (1-\alpha)d) - \max_{c \in A, e \in C} \kappa U(\alpha c + (1-\alpha)e) \end{aligned}$$

By linearity of U and V, preceding equals:

$$\begin{aligned} (1+\kappa)\left[\max_{a \in A} \alpha U(a) + \max_{c \in C}(1-\alpha)U(c)\right] - \max_{b \in A} \alpha V(a) + \max_{d \in C}(1-\alpha)V(c) \\ -\kappa\left[\max_{c \in A} \alpha U(c) + \max_{e \in C}(1-\alpha)U(e)\right] \end{aligned}$$

which is equal to $\alpha W(A) + (1-\alpha)W(C)$. Clearly, $W(A) \geq W(B) \Rightarrow \alpha W(A) + (1-\alpha)W(C) \geq \alpha W(B) + (1-\alpha)W(C)$, which implies $W(\alpha A + (1-\alpha)C) \geq W(\alpha B + (1-\alpha)C)$, and thus *Independence* holds.

To show *Partial Set-Betweenness*, note that it consists of two parts.

1. $A \succsim B \Rightarrow A \succsim A \cup B$

2. $A \succsim B$ *and* $A, B \in \mathcal{M}_c \Rightarrow A \cup B \succsim B$

Each element will be shown in turn. Begin with 1. Consider A, B s.t. $A \succsim B$. This implies $W(A) \geq W(B)$.

$$\begin{aligned} W(A) &= \max_{a \in A} \left[(1+\kappa)U(a) + V(a)\right] - \max_{b \in A} V(b) - \kappa \max_{c \in A} U(c) \\ W(B) &= \max_{a' \in B} \left[(1+\kappa)U(a') + V(a')\right] - \max_{b' \in B} V(b') - \kappa \max_{c' \in B} U(c') \\ W(A \cup B) &= \max_{a'' \in \{a, a'\}} \left[(1+\kappa)U(a'') + V(a'')\right] - \max_{b'' \in \{b, b'\}} V(b'') - \kappa \max_{c'' \in \{c, c'\}} U(c'') \end{aligned}$$

If $\max_{a \in A} \left[(1+\kappa)U(a) + V(a)\right] \geq \max_{a' \in B} \left[(1+\kappa)U(a') + V(a')\right]$, then clearly $W(A) \geq$

$W(A \cup B)$. If $\max_{a' \in B} [(1+\kappa)U(a') + V(a')] > \max_{a \in A} [(1+\kappa)U(a) + V(a)]$, then $W(B) > W(A \cup B)$, so again, clearly $W(A) \geq W(A \cup B)$.

Turn to 2. Given $A \succsim B$ and $A, B \in \mathcal{M}_c$, $\kappa \max_{c \in A} U(c) = \kappa \max_{c' \in B} U(c')$. Call this value γ. Thus,

$$\begin{aligned} W(A) &= \max_{a \in A} [(1+\kappa)U(a) + V(a)] - \max_{b \in A} V(b) - \gamma \\ W(B) &= \max_{a' \in B} [(1+\kappa)U(a') + V(a')] - \max_{b' \in B} V(b') - \gamma \\ W(A \cup B) &= \max_{a'' \in \{a,a'\}} [(1+\kappa)U(a'') + V(a'')] - \max_{b'' \in \{b,b'\}} V(b'') - \gamma \end{aligned}$$

If $\max_{a \in A} [(1+\kappa)U(a) + V(a)] \geq \max_{a' \in B} [(1+\kappa)U(a') + V(a')]$, then clearly $W(A \cup B) \geq W(B)$. If $\max_{a' \in B} [(1+\kappa)U(a') + V(a')] > \max_{a \in A} [(1+\kappa)U(a) + V(a)]$, then $W(A \cup B) = W(B)$, because the maxima are the same for each term.

Finally, to show *Singleton Set-Betweenness*, observe that

$$W(\{a\}) = U(a) = \max_{x \in a} [u(a) + v(a)] - \max_{y \in a} v(y);$$

by the Gul-Pesendorfer representation theorem, this satisfies *Set-Betweenness*.

Bibliography

Barbos, A. (2010). Context effects: A representation of choices from categories. *Journal of Economic Theory 145*, 1224–1243.

de Clippel, G. and K. Eliaz (2012). Reason-based choice: A bargaining rationale for the attraction and compromise effects. *Theoretical Economics 7*, 125–162.

Dekel, E., B. L. Lipman, and A. Rustichini (2001, July). Representing preferences with a unique subjective state space. *Econometrica 69*(4), 891–934.

Gul, F. and W. Pesendorfer (2001, November). Temptation and self-control. *Econometrica 69*(6), 1403–1435.

Herstein, I. and J. Milnor (1953, April). An axiomatic approach to measurable utility. *Econometrica 21*(2), 291–297.

Huber, J., J. W. Payne, and C. Puto (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research 9*, 90–98.

Kivetz, R., O. Netzer, and V. Srinivasan (2004). Alternative models for capturing the compromise effect. *Journal of Marketing Research 41*, 237–257.

Kopylov, I. (2007). A model of guilt and temptation. Working Paper.

Kopylov, I. (2009a). Finite additive utility representations for preferences over menus. *Journal of Economic Theory 144*, 354–374.

I use a result from the earlier working paper version, accessed at https://webfiles.uci.edu/ikopylov/www/files/fa_representations.pdf

Kopylov, I. (2009b). Temptations in general settings. *The B.E. Journal of Theoretical Economics 9*(1), Article 31.

List, J. A. and J. F. Shogren (1998, December). The deadweight loss of christmas: Comment. *American Economic Review 88*(5), 1350–1355.

Noor, J. and L. Ren (2015, March). Temptation and guilt. Working Paper.

Ok, E., P. Ortoleva, and G. Riella (2011). Theory of product differentiation in the presence of the attraction effect. Mimeo, New York University.

Ok, E., P. Ortoleva, and G. Riella (2015, January). Revealed (p)reference theory. *American Economic Review 105*(1), 299–321.

Pan, Y., S. O'Curry, and R. Pitts (1995). The attraction effect and political choice in two elections. *Journal of Consumer Pyschology 4*(1), 85–101.

Poterack, A. and B. Solow (2015, November). Extremist politics and the preference for compromise. Mimeo, Boston University.

Ravid, D. (2015, August). Focus, then compare. Mimeo, University of Chicago.

Rubinstein, A. and Y. Salant (2008). (a,f), choice with frames. *Review of Economic Studies 75*, 1287–1296.

Simonson, I. (1989). Choice based on reasons: The case for the attraction and compromise effects. *Journal of Consumer Research 16*, 158–174.

Simonson, I. and A. Tversky (1993). Context-dependent preferences. *Management Science 39*, 1179–1189.

Solnick, S. J. and D. Hemenway (1996, December). The deadweight loss of christmas: Comment. *American Economic Review 86*(5), 1299–1305.

Solow, J. L. (1993). Is it really the thought that counts? toward a rational theory of christmas. *Rationality and Society 5*(4), 506–517.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science 4*(3), 199–214.

Tserenjigmid, G. (2015, August). Choosing with the worst in mind: A reference-dependent model. Mimeo, California Institute of Technology. First draft 2014, current version 2015.

Tversky, A. and D. Kahneman (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics 106*(4), 1039–1061.

Waldfogel, J. (1993, December). The deadweight loss of christmas. *American Economic Review 83*(5), 1328–1336.

Alex Poterack

270 Bay State Rd, Room B02
Boston University, Department of Economics
Boston, MA, 02215 USA
Cell: (602) 334-7392
Fax: (617) 353-4449
Email: poterack@bu.edu
Website: http://poterack.weebly.com

Education

Ph.D. in Economics, Boston University, Boston MA, 2016
Dissertation Title: *Compromise, Extremism, and Guilt*
Main Advisor: Barton Lipman

B.A., Behavioral Economics and Public Policy (*Cum Laude*), Gallatin School of Individualized Study,
New York University, New York, NY, 2010

Fields of Interest

Behavioral Economics, Decision Theory, Health Economics, Industrial Organization

Teaching Experience

Instructor, International Trade, MET College, Boston University, Spring 2016

Instructor, Intermediate Microeconomic Analysis, Department of Economics, Boston University, Fall 2015, Spring 2016

Instructor, Introduction to Health Economics, Department of Economics, Boston University, Summer 2015

Instructor, International Trade, Department of Economics, Boston University, Summers 2013, 2014, 2015

Head Teaching Fellow, Introductory Macroeconomic Analysis, Department of Economics, Boston University, Spring 2015

Teaching Fellow, Introductory Microeconomic Analysis, Falls 2011, 2012, 2013, 2014, Spring 2014

Work Experience

Research Assistant for Barton Lipman, Boston University, 2013
Research Assistant for Ching-To Albert Ma, Boston University, 2012

Professional Service:
Officer of the BU Graduate Economics Association
Co-Organizer of Decision Theory Reading Group

Computer Skills: R, Stata, Matlab, MS Office, and LaTeX

Citizenship: USA

Working Papers

"The Compromise and Attraction Effects Through Frame Preferences" (Job Market Paper)

"Extremist Politics and the Preference for Compromise" (with Benjamin L. Solow)

"Giving the Gift of Guilt Avoidance"

Work in Progress

"Data Collector, M.D.: Moral Hazard and the Difficulty of Applying Big Data to Medicine"

References

Professor Barton Lipman
Department of Economics
Boston University
270 Bay State Rd
Boston MA 02215 USA
Phone: +1-617-353-2995
Email: blipman@bu.edu

Professor Jawwad Noor
Department of Economics
Boston University
270 Bay State Rd
Boston MA 02215 USA
Phone: +1-617-353-4436
Email: jnoor@bu.edu

Professor Dilip Mookherjee
Department of Economics
Boston University
270 Bay State Rd
Boston MA 02215 USA
Phone: +1-617-353-4392
Email: dilipm@bu.edu